8-2012 12:00 AM

Structural and Bioinformatic Analysis of the b Subunit of F-ATP synthase

Ardeshir Goliaei
The University of Western Ontario

Supervisor
Dr. Stan Dunn
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

Structural and Bioinformatic Analysis of the *b* Subunit of F-ATP synthase

(Spine title: Structural and Bioinformatic analysis of the *b* Subunit)

(Thesis format: Integrated-Article)

by

Ardeshir Goliaei

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO

SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
_____	_____
Dr. Stan Dunn	Dr. Susan Koval
	_____
Supervisory Committee	Dr. David Edgell
_____	_____
Dr. Greg Gloor	Dr. David Haniford
_____	
Dr. Brian Shilton	

The thesis by

Ardeshir Goliaei

entitled:

Structural and Bioinformatic Analysis of the *b* Subunit of F-ATP synthase

is accepted in partial fulfillment of the
requirements for the degree of
Master of Science

Date_____ _____

Chair of the Thesis Examination Board

Abstract and keywords

F_1F_o ATP synthases are rotary enzymes that produce most ATP in living organisms. The enzymes' $b_2\delta$ subunits form a stator stalk that holds the F_1 sector against the torque of the rotor. In *Escherichia coli*, b_2 is an asymmetric homodimer. However, ATPases from some species have a heterodimer of subunits *b* and *b'*. Here, a modified *E. coli* ATP synthase containing a heterodimeric stator stalk was engineered by replacing residues 34-110 of *E. coli b* with sequences of *Rhodobacter capsulatus b* and *b'*, and expressing both chimeras. This produced a functional, heterodimeric enzyme useful for investigating the two *b* subunits through mutation. Then we studied the evolution of *b* and *b'*. The heterodimeric system has evolved at least four times, and the gene duplication is usually linked with rearrangement of ATP synthase genes into multiple transcriptional units. Groups with each gene pattern mapped as clades onto a tree of life.

Keywords: ATP synthase, *b* subunit, stator stalk, right-handed coiled-coil, hendecad, disulfide formation, ATP synthase purification, *b/b'* evolution, homology, ribosomal protein phylogeny, tree of life, ATP synthase gene structure

Acknowledgment

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Stan Dunn who has started to teach me from the very first moment we met. His patience, motivation, enthusiasm, and immense knowledge has been a source of continuous support for me throughout my study. Other than the scientific aspects, he has been a great example of positive influence in my regular life. I could not have imagined having a better supervisor and mentor for my MSc study.

I would like to thank my advisory committee members, Dr. Greg Gloor and Dr. Brian Shilton for their guidance and input. Specifically Dr. Gloor whose innovative ideas and suggestions has always helped me stay focused and concentrated.

I would also like to thank my collaborator Russel Dickson for all of his help and work throughout this project. He was always ready to explain difficult concepts in a simple and understandable way. His constant help was extremely valuable to me.

Many thanks to Yumin Bi, our lab technician whose technical assistance has made performing experiments possible. With her support, working in the lab was always fun.

Lastly, and most importantly, I am grateful to my beloved wife Mahshid Ashkzari, for her continuous support and help in my life. Her love, support, and encouragement was what made work and research possible for me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations, Acronyms and Short Forms

ADP .....adenosine diphosphate

Amino acids :

A	ala	alanine
C	cys	cysteine
D	asp	aspartic acid
E	glu	glutamic acid
F	phe	phenylalanine
G	gly	glycine
H	his	histidine
I	ile	isoleucine
K	lys	lysine
L	leu	leucine
M	met	methionine
N	asn	asparagine
P	pro	proline
Q	gln	glutamine
R	arg	arginine
S	ser	serine
T	thr	threonine
V	val	valine
W	trp	tryptophan

Y tyr tyrosine

ACMA 9-amino-6-chloro-2-methoxyacridine

ATP .....adenosine triphosphate

b^C .....in an offset dimer, the *b* shifted C-terminally

b^N .....in an offset dimer, the *b* shifted N-terminally

BLAST.....basic local alignment search tool

BSA .....bovine serum albumin

CD .....circular dichroism

CDS.....coding sequences

CuCl2 .....copper (II) chloride

DNA .....deoxyribonucleic acid

DTT .....dithiothreitol

EDTA .....ethylenediamine tetra-acetic acid

HCl .....hydrochloric acid

LB .....Luria broth

LHCC .....left-handed coiled coil

$MgCl_2$ .....magnesium chloride

NaCl .....sodium chloride

NCBI.....national center for biotechnology information

NEM .....N-ethyl maleimide

P_i .....inorganic phosphate

PMSF .....phenylmethylsulfonyl fluoride

PSI-BLAST.....position-specific iterative basic local alignment search tool

PSSM .....position-specific scoring matrix

RHCC .....right-handed coiled coil

SDS-PAGE .....sodium dodecyl sulfate polyacrylamide gel electrophoresis

TCEP-HCl .....tris (2-carboxyethyl)-phosphine hydrochloride

Tris .....tris (hydroxymethyl) aminomethane

UV .....ultraviolet

YT .....yeast extract tryptone

Chapter 1: General introduction

1.1 ATP synthase

1.1.1 Oxidative phosphorylation and role of ATP synthase.

Living organisms need a continuous source of energy like light or nutrients to support their energy demands. To be able to use these external sources cells need a system to convert them to the universal energy currency, adenosine triphosphate (ATP). Oxidative phosphorylation is the system in which energy released from oxidation of nutrients is used to produce ATP. Release of energy by hydrolysis of ATP to ADP and phosphate drives a large number of energy requiring processes. The resultant ADP and phosphate must then be continuously converted back to ATP. Every day this process leads to the turnover of 50 kg of ATP in a human body on average.

One major set of reactions that are performed during oxidative phosphorylation are redox reactions in which electrons are transferred through electron transport chains to a terminal electron acceptor, most often oxygen. The membrane-embedded protein complexes that catalyse these reactions utilize the energy released by the redox reactions to drive transport of protons across the inner mitochondrial membrane in eukaryotes, or cytoplasmic membranes in prokaryotes. Since these coupling membranes are intrinsically impermeable to protons, this process generates an electrochemical potential across the membrane. Flow of protons back across the membrane and down the generated gradient occurs through a large enzyme complex called ATP synthase. The ATP synthase is a rotary mechanical motor and as a result of proton flow, parts of the complex rotate

relative to the rest, leading to the generation of ATP from ADP and phosphate at the catalytic sites. The translocation of protons by respiratory components and by ATP synthase, with the use of the proton gradient as an intermediate between the two was envisaged and given the name “chemiosmosis” by Peter Mitchell (1,2).

1.1.2 Different types of ATP synthases.

ATP synthases are membrane bound molecular motors which can have two different functions depending on their type and the organism they originate from (Figure 1.1). In some cases they work as an ATP driven proton pump to acidify interior parts of cellular components and in other cases utilize the potential gradient across the membrane to generate ATP out of ADP and phosphate. In general they have been divided into three major groups based on their phylogenetic origins: A-ATPases, found in *Archaea* and some *Bacteria,* involved in ATP synthesis; F-ATPases, found in inner membrane of mitochondria, chloroplasts, and bacteria and their function is also ATP synthesis; and V-ATPases, found in eukaryotes specifically in vacuoles and vesicles where they acidify the inner space of these compartments (3,4,5). These three types of ATPases are related to each other and there are sequence, structural, and functional similarities among them (6,7). A and V types especially are very similar and some *Bacteria* have ATPase of type A/V, acquired through lateral gene transfer (8).

1.1.3 The F-type ATP synthase and the two sectors F_1 and F_o.

F_1F_o ATP synthases are mushroom-like multisubunit complexes that are composed of two structurally and functionally distinct parts; the membrane-peripheral F_1 sector

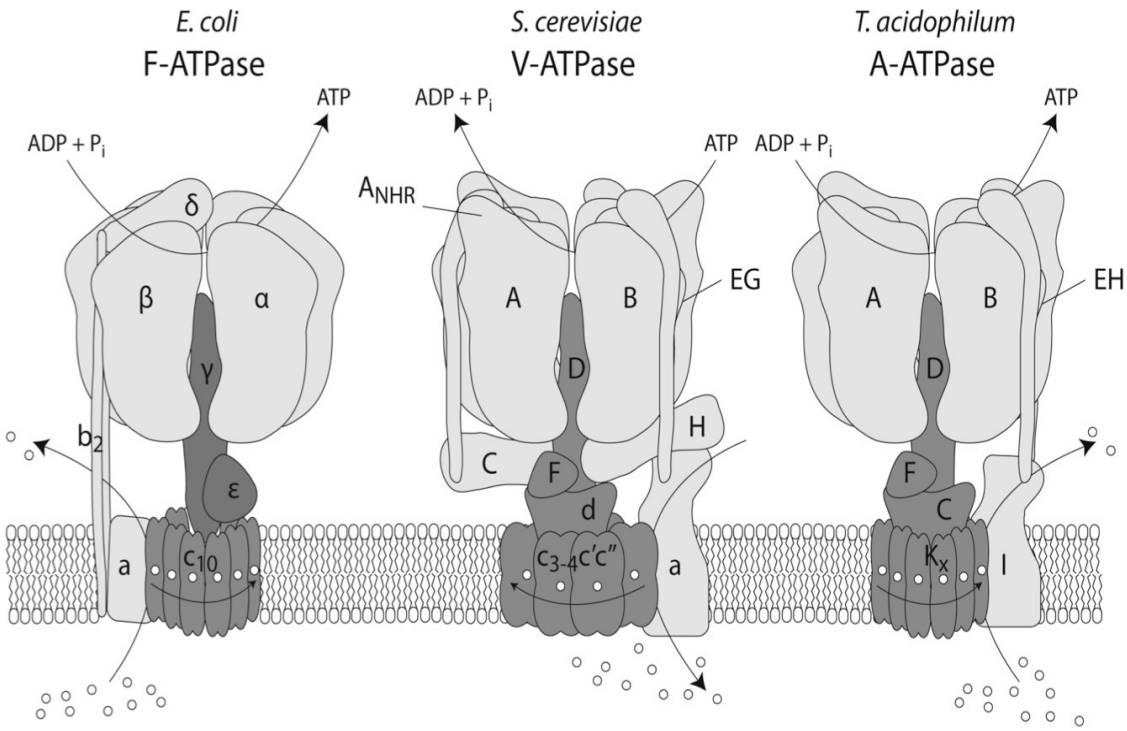

Figure 1.1: Different types of ATPases. These are three classes of rotary ion translocating ATPases. They are all related to each other, there is sequence and structure similarity among these types, but there are differences between the three classes as well. F-ATPases are found in mitochondria, chloroplasts, and bacteria and its main function is to use the proton gradient across the membrane for ATP synthesis. A-ATPases have the same function but are found in archaea. V-ATPases are found in eukaryotes specifically in vacuoles and vesicles where they acidify the inner space of these compartments. Figure reproduced from reference (51) with permission.

(subunits $\alpha_3\beta_3\gamma\delta\varepsilon$) and the membrane-integral F_o sector (subunits ab_2c_{10} in *Escherichia coli*) (Figure 1.2). The enzyme works by rotation of a subset of subunits against the other; the rotor ($\gamma\varepsilon c_{10}$) against the stator ($\alpha_3\beta_3\delta ab_2$) (9). This rotation happens as the result of the movement of ions downhill according to their gradient. The mechanism of coupling ion translocation across the F_o to the rotation of the *c* ring is not completely understood. However, it is believed to happen by two independent, disconnected half channels that are present in the interface of the *a* subunit and the *c* ring. While the *c* ring rotates each *c* subunit in the ring is protonated. Rotation brings each one of the protonated *c* subunits into the first half channel where they release their proton into the cell interior (cytoplasm). Then, when the deprotonated *c* subunit reaches the other half channel, protons from outside (periplasm) that have entered this half channel can bind to the *c* subunit which then continues to rotate around the ring and goes through another cycle. In *E. coli*, the guanidine group of a conserved arginine residue in the *a* subunit, Arg-212, is believed to play a major role in protonation and deprotonation of the *c* subunits (10,11).

The two sectors are connected by two stalks (Figure 1.2). The central stalk contains the γ and ε subunits and it protrudes into the $\alpha_3\beta_3$ assembly in F_1. The rotation of asymmetric γ subunit drives a cycle of conformational changes in the catalytic β subunits, finally resulting in ATP synthesis and release (12). The peripheral stalk, which extends from the membrane to the top of F_1, approximately 125 Å from the membrane, is formed by association of a b_2 dimer with the δ subunit. By preventing $\alpha_3\beta_3$ complex from rotating with γε, the peripheral stalk serves as a stator so that the conformational changes in the

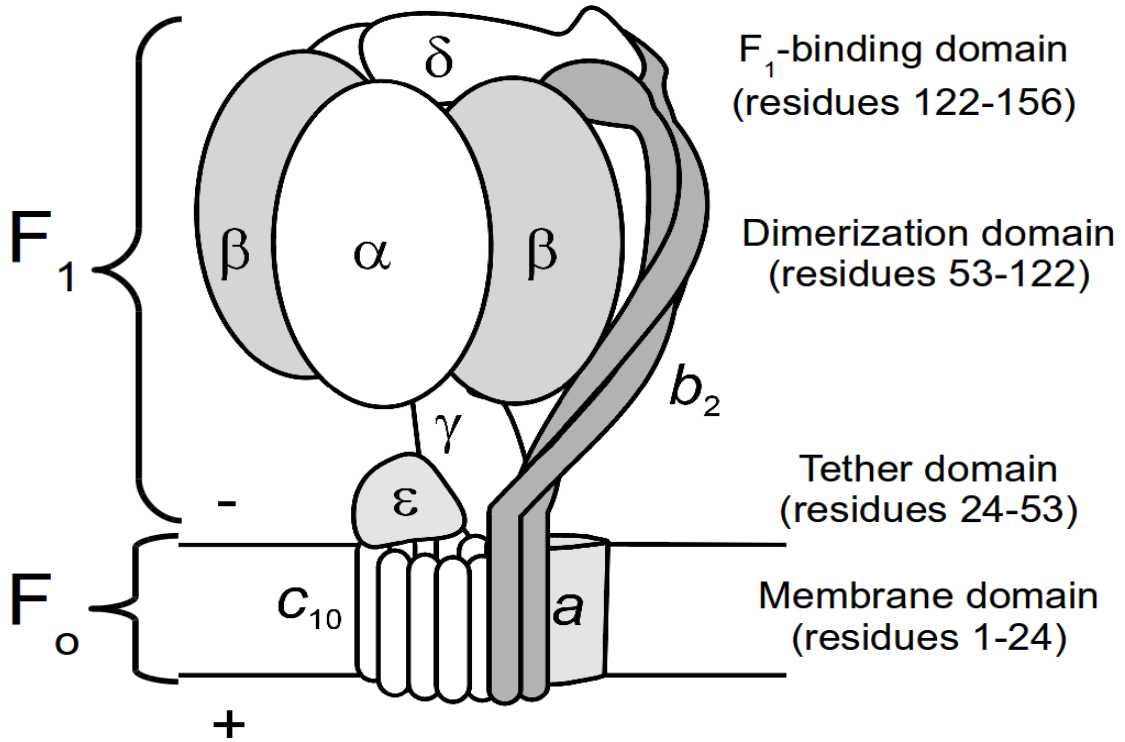

Figure 1.2: Schematic structure of *E. coli* F-type ATP synthase and the *b* subunit with its different domains. F_1F_o ATP synthases are composed of two linked sectors: the membrane-integral F_o sector which catalyzes the movement of protons across the membrane and the membrane-peripheral F_1 sector which catalyzes the synthesis or hydrolysis of ATP. The F_1 sector of the prototypical ATP synthase from *Escherichia coli* consists of five different polypeptide chains in a stoichiometry of $\alpha_3\beta_3\gamma\delta\varepsilon$; the F_o from this bacterium is composed of three subunits in a stoichiometry of ab_2c_{10}. The rotor part, $c_{10}\gamma\varepsilon$, rotates as protons translocate through F_o channel and this rotation induces conformational shifts, necessary for synthesis of ATP, in the catalytic sites in the β subunits. The enzyme's peripheral stalk, composed of the $b_2\delta$ subunits, serves as the stator that holds the F_1 sector and its catalytic sites against the movement of the rotor. The *b* subunit of peripheral stalk in *E. coli* consists of 156 residues and has a molecular weight of 17264 Da. It has been divided into four domains. The insoluble N-terminal part of *b* subunit which resides in the membrane, is called transmembrane domain (b_{M1-I24}). The soluble portion consists of the tether domain ($b_{E24-A53}$) which is the segment between the surface of the membrane and the bottom of F_1 sector, the dimerization domain ($b_{T53-K122}$) which extends up to one of the αβ interfaces in F_1, and at the end, the F_1-binding domain which is the C-terminal of *b* subunit and is required for the interaction of the peripheral stalk and F_1 ($b_{Q123-L156}$).

three β subunits can be induced during rotational catalysis (13). Other bacteria and chloroplasts have a similar architecture for their peripheral stalk but mitochondrial stalks are more complex and consist of four different subunits (13).

1.2 The *b* subunit of *E. coli*

1.2.1 Sequence

The *b* subunit of ATP synthase in *E. coli* has 156 residues and it is a 17,264 Da polypeptide. Sequence analysis of this protein reveals two distinct regions in the protein; a hydrophobic N-terminal membrane spanning region, and the remaining hydrophilic region which is rich in alanine and polar residues. Alignment of this protein with *b* subunits from other bacteria and chloroplasts shows them to contain the same features in general and they are easily aligned, especially in the hydrophilic region (Figure 1.3).

In *E. coli* and most other *Bacteria* there is only one gene for the *b* subunit and the resulting polypeptides form a homodimer. However, peripheral stalks in *Bacteria* are not always homodimeric. In some cases there are two genes present in the ATP synthase operon and encode for two different types of *b* subunits, called *b* and *b'*. The existence of these two different genes was first discovered in a number of species among *Cyanobacteria* (14,15). Chloroplasts also have two subunits in their stator stalks called subunits I and II which are related to the bacterial ones. By the increasing availability of the number of completely sequenced *Bacteria*, it has become possible to find more organisms with heterodimeric stator stalks. For example, based on the present study *Alphaproteobacteria, Epsilonproteobacteria, Deltaproteobacteria, Cyanobacteria,*

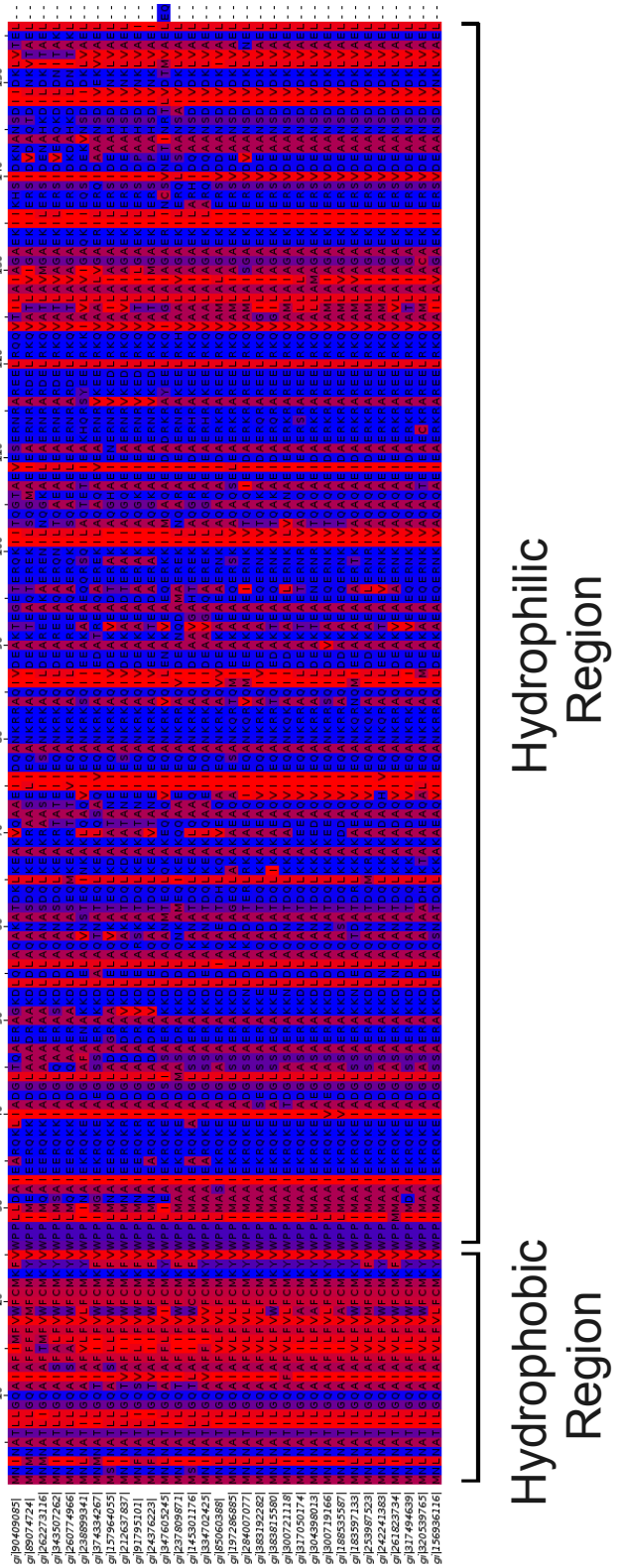

Figure 1.3: Multiple sequence alignment of a number of *b* subunits from different bacteria. Two distinct regions are identifiable: an N-terminal hydrophobic region and an hydrophilic segment rich in alanine. Hydrophobic residues are in “red” and hydrophilic residues are in “blue” color.

Actinobacteria, Acidobacteria, Deferribacteres, Aquificae, etc. are different taxa of *Bacteria* containing heterodimeric stator stalks. An interesting observation is that in *Mycobacteria* and some other members of the phylum of *Actinobacteria* the genes for b and δ are fused together to form a single gene but separated by a linker of about 100 residues. The *b'* in these organisms is encoded by a different open reading frame which precedes the bδ fused gene.

Based on several observations it is believed that the ATP synthase in heterodimeric organisms has one copy of each type of the *b* subunits. Purified ATP synthase from some species including *Rhodobacter capsulatus* (16) and *Aquifex aeolicus* (17) have been shown to contain both *b* and *b'* proteins. There are some studies that support the preferential heterodimerization of *b* and *b'*. Sedimentation and chemical cross-linking studies showed that the cytoplasmic domains of the *b* and *b′* subunits of *Synechocystis* form a highly extended, helical heterodimer. These data support the idea that *b/b'* heterodimers have structures similar to those of b_2 homodimers (18).

1.2.2 Structure

Up until now a crystal structure of the *b* subunit of *E. coli* or its entire cytoplasmic domain has not been solved, although a segment of the monomeric form of $b_{62\text{-}122}$ has been crystallized and shown to exist as a helix (19). Much of what we know about the structure of this protein is based on biochemical studies. Residues 25-156 of this protein has been expressed as a soluble dimer and shown to exist as an elongated homodimer. Circular dichroism studies revealed an α-helical coiled coil structure (20). By help of

deletion studies this structure has been further divided into four distinct domains (21) which will be discussed in more detail (Figure 1.2):

Membrane domain (residues 1-24). This region is the highly hydrophobic region of the *b* subunit and crosses the membrane once. An NMR structure of the $b_{1\text{-}34}$ construct has been determined, in monomeric form, which shows hydrophobic α-helix with a bend at residues 23-26. Also, disulfide formation of cysteines inserted in positions 2-20 were studied. The only substitution with intense high yield dimers were observed for N2C, T6C, Q10C, and F14C. This implies the pattern that is expected for one face of an α-helix interacting with the other α-helix (22).

Tether domain (residues 24-53). This domain is the region between transmembrane and dimerization domain (*vide infra*). It can be deleted from the soluble construct without any effect on dimerization (23). A surprising feature of this region is its tolerance for residue deletion or insertion. It has been shown that deletion of up to 11 residues (24) or insertion of up to 14 residues (25) can be accommodated in this region of the protein. Another feature of this region is the presence of a very conserved Arginine at position 36. Mutation of this residue has been shown to reduce the functionality of the enzyme by disrupting the assembly and coupling (26). This residue also has been shown to be close to the *a* subunit through chemical cross-linking studies. Based on this it appears that parts of tether domain are in contact with the cytoplasmic segment of the *a* subunit (27).

Dimerization domain (residue 53-122). The isolated dimerization domain of the *b* subunit has been characterized as an atypical, parallel, two-stranded coiled coil which is

necessary for interaction with F_1(28). A crystal structure of a large portion of the dimerization domain has been solved in monomeric form ($b_{62\text{-}122}$) and shows an amphipathic helix with right-handed pattern of hydrophobic residues along one side of the helix (19). Sequence analyses of the dimerization domain also have identified an 11-residue hendecad pattern, with positions denoted *abcdefghijk* (28) (Figure 1.4, A). Hendecad patterns are indicative of right-handed coiled coils in which 11 residues make three turns of the helix relative to the interhelical axis (29). The expected distribution of positions for a two-stranded structure of this type is shown on the helical wheel in Figure 1.4, B. Del Rizzo *et al.* (30) confirmed the assignment of *abcdefghijk* by introducing cysteine mutations in these positions and assessment of the stabilities of disulfide-linked dimers. Based on these results positions *a, d, e,* and *h* are the dimerization interface in the isolated dimerization domain where *a* and *h* positions at the centre of the interface are usually occupied by alanine or other small amino acids, whereas larger hydrophobic residues are often seen in the *d* and *e* positions that are more peripherally situated. Also, denaturation studies revealed that offset dimers (with disulfides between positions *a* and *h*) show greater thermal stability than in-register ones and their shapes are similar to that of the wild-type uncross-linked dimers (in-register dimers caused perturbation of the shape compared to wild-type). These results implied that the two *b* subunits are offset by 5.5 residues in the dimerization domain and one of the subunits is shifted toward the N-terminal and the other toward the C-terminal end of the dimer. The C-terminally shifted subunit was labeled b^C and the other subunit was called b^N. This observation indicates

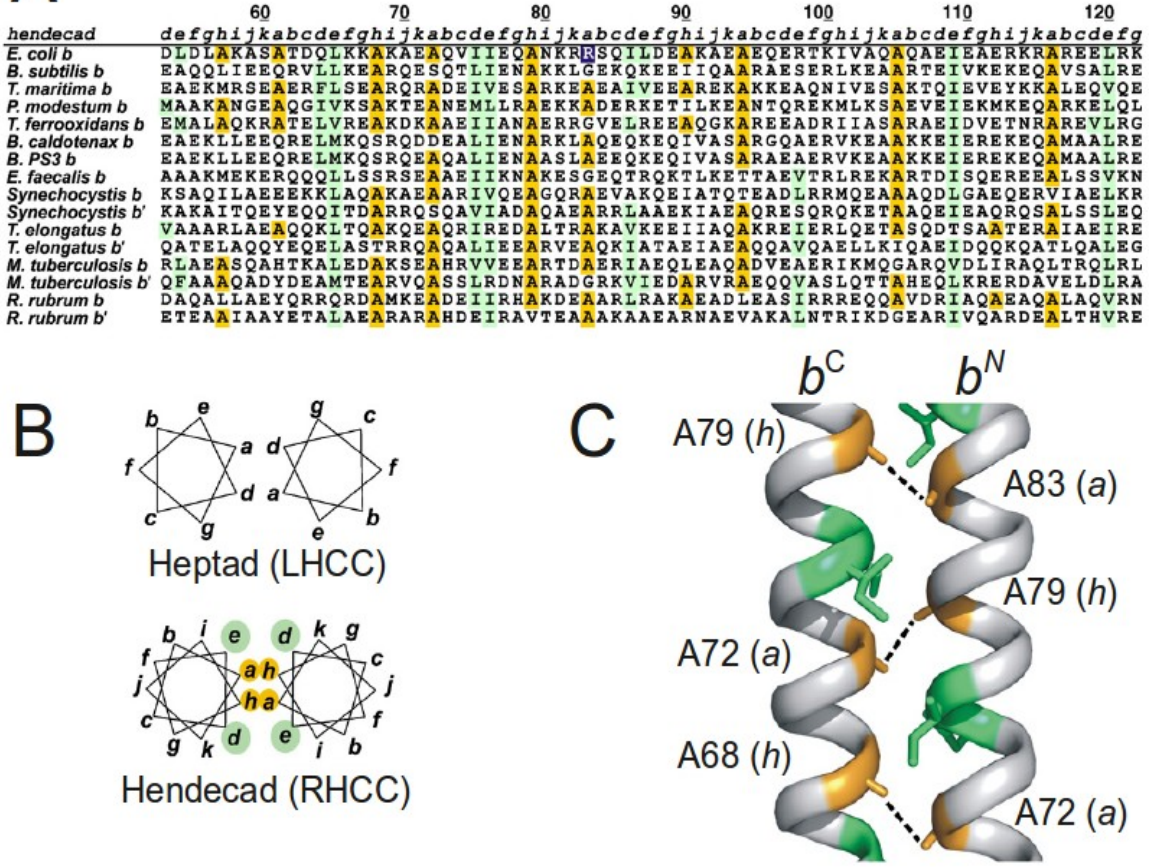

Figure 1.4: Sequence analysis of the dimerization domain of the *b* subunit. *Panel A,* multiple sequence alignment of the dimerization domain of several bacteria which has identified an 11-residue hendecad pattern, with positions denoted *abcdefghijk*. Hendecad patterns are indicative of right-handed coiled coils (RHCC) in which 11 residues make three turns of the helix relative to the interhelical axis. *Panel B,* the expected distribution of positions for a two-stranded structure of this type is shown on the helical wheel. The *a* and *h* positions at the centre of the interface are usually occupied by alanine or other small amino acids, whereas larger hydrophobic residues are often seen in the *d* and *e* positions that are more peripherally situated. In comparison, in a left-handed coiled-coil (LHCC) structure positions *d* and *e* are at the interface of the two helices. *Panel C,* the proposed model based on experimental data a RHCC structure in which the two *b* subunits are offset by 5.5 residues in the dimerization domain and one of the subunits is N-terminally shifted and the other one is shifted toward the C-terminal. The C-terminally shifted subunit was labelled b^C and the other one b^N. Figure reproduced from reference (30) with permission.

that the homodimer is intrinsically asymmetric.

F_1-Binding domain (residues 123-156). This region is essential for the binding of F_1 through δ subunit and has been suggested to have a more globular structure (31). Circular dichroism indicates helical structure for this domain but no well-defined heptad or hendecad pattern is observed in this region (23).

Since there is no dimeric high-resolution structure available for the *b* subunit, a model has been proposed based on experimental data explained above (Figure 1.4, C). In this model the two helices are arranged in a right-handed coiled coil structure based on sequence analysis, the crystal structure of the monomeric *b* subunit, and cysteine cross-linking studies in the dimerization domain. The two helices are also offset with respect to one another in contrast to in-register left-handed coiled coils. Recently, the crystal structure of a related stator stalk of A-type ATP synthase of *Thermus thermophilus* has been solved (32). This EG heterodimer contains a right-handed coiled-coil in which the E helix is virtually straight while the G subunit bends around it. This study not only has supported the unusual right-handed coiled coil model of the *b* subunit in F-type ATP synthases but also has extended the prediction of this kind of structure to other types of H^+-ATPases.

1.2.3 Interactions

Interactions of the *b* subunit with the reset of ATP synthase can be studied in three different regions:

1) F_o interactions. The idea of direct interactions between subunits *a* and *b* was

first indicated in 1986 by the finding that the mutation of a proline at position 240 to alanine or leucine suppressed the effects of mutation of residue 9 in the *b* subunit (33). Also, strong interaction between subunits *a* and *b* has been shown by the purification of a stable ab_2 subcomplex which reflects subunit interactions occurring within the F_o complex *in vivo* (34). Cysteine cross-linking studies also revealed strong disulfide formation between cysteines inserted at positions 227 and 228 in the *a* subunit and position 2 in the *b* subunit (35). Whether both of the two copies of *b* interact individually with the subunit *a* or whether just one of the *b* subunits is in direct contact with *a* remains to be elucidated. It was also previously mentioned that cysteine introduced in position 36 of the *b* subunit can be cross-linked to subunit *a* which shows another point of interaction between the *a* and *b* subunits (27).

Interaction of the *b* and *c* subunits was also first indicated by observing that mutation of Alanine-62 of subunit *c* to serine suppressed the effects of a mutation in the position 9 of the *b* (36). Disulfide formation studies revealed that cysteine introduced in position 2 of the subunit *b* can be cross-linked with cysteine at positions 74, 75, and 78 of the *c* subunit. In each case a maximum of 50% of the *b* subunit could be cross-linked to subunit *c* (37). This observation suggests that only one of the two *b* subunits lies proximal to the *c*-ring of F_o and the second *b* may lie more to the periphery of the complex, perhaps in association with subunit *a*.

2) F_1 interactions. Results of cross-linking experiments has provided evidence for the interaction of the *b* subunit with the surface of F_1, more specifically α and β subunits.

Based on these results position 92 of the *b* subunit was found to cross-link to α (between residues 464 and 483) and a nearby position in β near the bottom of $\alpha_3\beta_3$ complex at a noncatalytic interface. Also position 109 was found to cross-link in an upper position in *α* (between residues 213 and 220) (27).

Binding of the *b* subunit to the top of F_1 is mediated through formation of $b_2\delta$ complex which has been demonstrated by several studies. C-terminal truncation or proteolysis experiments provided strong evidence that the C-terminal regions of both *b* and δ proteins are involved in formation of $b_2\delta$ complex (38,39,40,41). The proposed offset model in the dimerization domain with the two helices being shifted and asymmetric with respect to each other (30) has raised the question of whether both C-terminals of the two *b* subunits are involved in the interaction with δ or only one of them. Experiments with different disulfide-linked dimers in which subunit positions were fixed by formation of an asymmetric disulfide bond has revealed that offset dimers, compared to in-register homodimers, bind more tightly to F_1 and that only b^N could be cross-linked to δ. Also, these studies showed that b^N is more important for F_1 binding than b^C (42). These results imply that the two *b* subunits are not only asymmetric with respect to each other but also have different interactions with the ATP synthase complex and play different roles in the peripheral stalk.

3) No interaction area. Based on cross-linking and mass spectrometry studies it is known that the section of the *b* subunit in between positions 55-91 is located between F_o and the lower surface of F_1 where there is no contact with other subunits of ATP synthase.

This area is the section in which right-handed coiled-coil structure has been established by using disulfide formation of *a* and *h* positions between the two subunits. Lack of any specific tight contacts with the rest of the enzyme also has been shown by observing that replacement of this region by the corresponding sequences from other bacterial *b* subunits does not have any effect on assembly or function (43). This region has been used as a logical candidate for making a heterodimeric system in *E. coli* using chimeric approaches (44,45).

1.2.4 Function in the stator stalk

As the *b* subunit is the most extended part of the stator stalk its main function is to anchor catalytic site to the membrane and hold F_1 and F_o together and to prevent the $\alpha_3\beta_3$ complex from rotating with the rotor. This suggests a rigid and strong structure for the stator, strong enough to be able to withstand the torque. However, since the rotation of F_1 and F_o are stepwise and not fine-tuned to one another elastic power transmission has been claimed to exist between the two (46,47,48,49). More specifically, in *E. coli*, F_o has a ring of 10 *c* subunits and rotates in ten steps, while F_1 has three αβ dimers and the γ subunit rotates in three steps. Accordingly, production of 3 ATP molecules per 10 protons pumped across the membrane implies variable stoichiometry for production of ATP (production of some ATP molecules needs 3 protons and some needs 4). This variable stoichiometry necessitates elastic linkages between the two sectors. Recently, single molecule studies have found the major elastic buffer between F_1 and F_o (50). Based on these results the contact region between subunits γ and ε with the c_{10} ring (lower portion

of the rotor) is a compliant domain while the stator, with the dimer of b_2 as its most extended component, is stiffer than the rotor by at least one order of magnitude. When the stiffness of the wild type stator was lowered by using two mutants (in one mutant length of the *b* subunit was increased and in the other mutant it was destabilized by introducing three glycine mutations) the functionality of the enzyme decreased (as observed by functionality of the enzyme by growth on succinate media and measuring proton pumping activity). These observations showed that the stator is stiff and it has to be stiff for the enzyme to work. In the mentioned study glycine mutations were introduced to both of the *b* subunits at the same time since they are homodimer of identical proteins in *E. coli*. It is interesting to know if the two *b* subunits contribute to this stiffness equally since the recent crystal structure of the related A-type ATP synthase stator stalk shows a difference in the structure of the two helices forming the dimer; one of the helices is almost straight while the other one has a lot of glycine residues in its sequence and bends around the straight one (32) (Figure 1.5).

1.3 Focus of this study

In the absence of a complete high-resolution structure for the *b* dimer and in fact of the ATP synthase there is still a lot to be learned regarding its structure, interactions, and function. These studies have been challenging by the fact that ATP synthase in *E. coli* has a homodimeric stator stalk and this limits our ability to manipulate the two subunits independently. We know that these polypeptides have different interactions with the ATP synthase enzyme and occupy different positions relative to one another. In fact

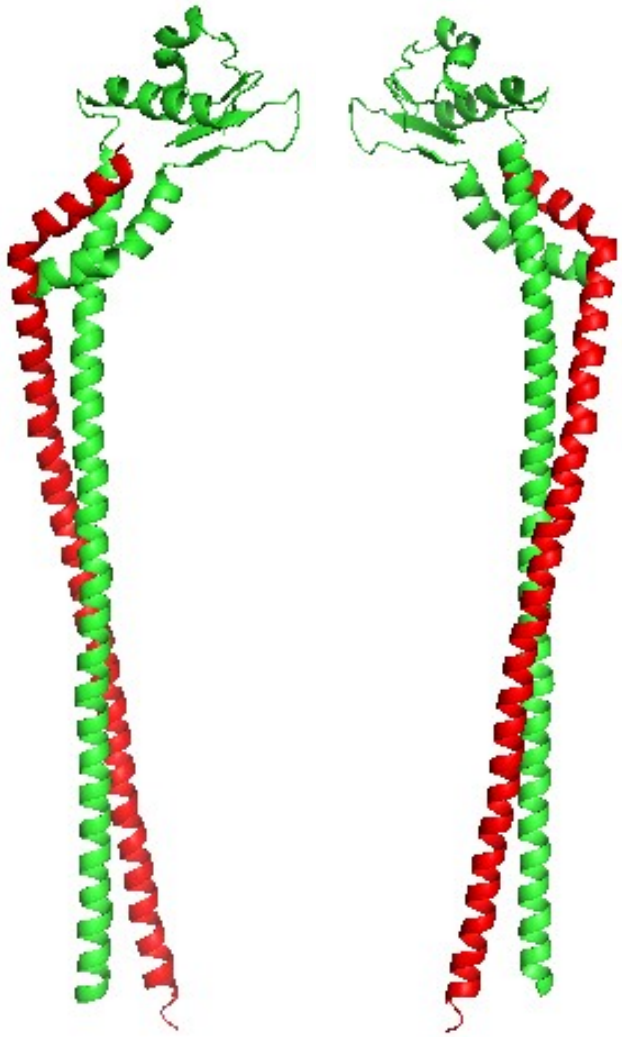

Figure 1.5: Structure of the stator stalk of A-type ATP synthase of *Thermus thermophilus* (pdb:3K5B). Two different views of the structure of the EG stator stalk of the A-type ATP synthase of Thermus thermophilus are shown in the figure. Subunit E is coloured green and subunit G is coloured red. The views are rotated 180 degree relative to each other. The two subunits form a right-handed coiled-coil structure in the N-terminal region and a globular structure in the C-terminal part. The E subunit is more straight and the G subunit bends around it to form a coiled-coil structure.

evolutionary development of heterodimeric stator stalks in other bacteria or chloroplasts seems likely to have happened for the purpose of specialization of the two proteins making one of them more efficient for the b^N and the other one for the b^C.

Chapter 2 of this thesis is about the design, development, and testing of a system that allows the easy and independent manipulation of each one of the *b* subunits in *E. coli*. This makes it possible to design experiments for determining their specific interactions with the rest of ATP synthase subunits. We have used this system to investigate some of the unanswered questions regarding the structure of the *b* subunit, like the offset model in the transmembrane domain, or specific role of each one of the *b* subunits in the stiffness of the stator.

In Chapter 3 the evolution of heterodimeric systems has been studied using bioinformatics tools to see if specialization of b^N and b^C subunits can be rationalized by identification of relevant sequence elements. Phylogenetic analysis of heterodimers accompanied by knowledge about their ATP synthase gene structure has provided a promising way of classifying these organisms to further deepen our understanding of the evolution of heterodimeric stator stalks.

1.4 References

1 Mitchell, P. (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature.* **191**, 144-8.

2 Mitchell P. (1966) Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biol Rev Camb Philos Soc.* **41**, 445-502.

3 Harvey WR. (1992) Physiology of V-ATPases. *J Exp Biol.* **172**, 1-17.

4 Nelson N. (1989) Structure, molecular genetics, and evolution of vacuolar H+-ATPases. *J Bioenerg Biomembr.* **21**, 553-71.

5 Schäfer G, Meyering-Vos M. (1992) F-type or V-type? The chimeric nature of the archaebacterial ATP synthase. *Biochim Biophys Acta.* **1101**, 232-5.

6 Imamura H, Nakano M, Noji H, Muneyuki E, Ohkuma S, Yoshida M, Yokoyama K. (2003) Evidence for rotation of V1-ATPase. *Proc Natl Acad Sci U S A.* **100**, 2312-5.

7 Zimniak L, Dittrich P, Gogarten JP, Kibak H, Taiz L. (1988) The cDNA sequence of the 69-kDa subunit of the carrot vacuolar H+-ATPase. Homology to the beta-chain of F0F1-ATPases. *J Biol Chem.* **263**, 9102-12.

8 Olendzenski L, Liu L, Zhaxybayeva O, Murphey R, Shin DG, Gogarten JP. (2000) Horizontal transfer of archaeal genes into the deinococcaceae: detection by molecular and computer-based approaches. *J Mol Evol.* **51**, 587-99.

9 Noji H, Yasuda R, Yoshida M, Kinosita K Jr. (1997) Direct observation of the rotation of F1-ATPase. *Nature.* **386**, 299-302.

10 Vik SB, Antonio BJ. (1994) A mechanism of proton translocation by F1F0 ATP synthases suggested by double mutants of the a subunit. *J Biol Chem.* **269**, 30364-9.

11 Junge W, Lill H, Engelbrecht S. (1997) ATP synthase: an electrochemical transducer with rotatory mechanics. Trends Biochem Sci. **22**, 420-3.

12 Boyer PD. (1993) The binding change mechanism for ATP synthase--some probabilities and possibilities. *Biochim Biophys Acta.* **1140**, 215-50.

13 Dunn SD, McLachlin DT, Revington M. (2000) The second stalk of Escherichia coli ATP synthase. *Biochim Biophys Acta.* **1458**, 356-63.

14 Cozens AL, Walker JE. (1987) The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium Synechococcus 6301. Support for an endosymbiotic origin of chloroplasts. *J Mol Biol.* **194**, 359-83.

15 McCarn DF, Whitaker RA, Alam J, Vrba JM, Curtis SE. (1988) Genes encoding the alpha, gamma, delta, and four F0 subunits of ATP synthase constitute an operon in the cyanobacterium Anabaena sp. strain PCC 7120. *J Bacteriol.* **170**, 3448-58.

16 Gabellinia N, Gaoa Z, Eckerskorna C, Lottspeicha F, Oesterhelta D. (1988) Purification of the H+-ATPase from Rhodobacter capsulatus, identification of the F1F0 components and reconstitution of the active enzyme. *Biochim Biophys Acta - Bioenergetics.* **934**, 227-234.

17 Peng G, Bostina M, Radermacher M, Rais I, Karas M, Michel H. (2006) Biochemical and electron microscopic characterization of the F_1F_0 ATP synthase from the hyperthermophilic eubacterium Aquifex aeolicus. *FEBS Lett.* **580**, 5934-40.

18 Dunn SD, Kellner E, Lill H. (2001) Specific heterodimer formation by the cytoplasmic domains of the b and b' subunits of cyanobacterial ATP synthase. *Biochemistry.* **40**, 187-92.

19 Del Rizzo PA, Bi Y, Dunn SD, Shilton BH. (2002) The "second stalk" of Escherichia coli ATP synthase: structure of the isolated dimerization domain. *Biochemistry.* **41**, 6875-84.

20 Dunn SD. (1992) The polar domain of the b subunit of Escherichia coli F1F0-ATPase forms an elongated dimer that interacts with the F1 sector. *J Biol Chem.* **267**, 7630-6.

21 Revington M, McLachlin DT, Shaw GS, Dunn SD. (1999) The dimerization domain of the b subunit of the Escherichia coli F(1)F(0)-ATPase. *J Biol Chem.* **274**, 31094-101.

22 Dmitriev O, Jones PC, Jiang W, Fillingame RH. (1999) Structure of the membrane domain of subunit b of the Escherichia coli F0F1 ATP synthase. *J Biol Chem.* **274**, 15598-604.

23 Revington M, Dunn SD, Shaw GS. (2002) Folding and stability of the b subunit of the F(1)F(0) ATP synthase. *Protein Sci.* **11**, 1227-38.

24 Sorgen PL, Caviston TL, Perry RC, Cain BD. (1998) Deletions in the second stalk of F1F0-ATP synthase in Escherichia coli. *J Biol Chem.* **273**, 27873-8.

25 Sorgen PL, Bubb MR, Cain BD. (1999) Lengthening the second stalk of F(1)F(0) ATP synthase in Escherichia coli. *J Biol Chem.* **274**, 36261-6.

26 Caviston TL, Ketchum CJ, Sorgen PL, Nakamoto RK, Cain BD. (1998) Identification of an uncoupling mutation affecting the b subunit of F1F0 ATP synthase in Escherichia coli. *FEBS Lett.* **429**, 201-6.

27 McLachlin DT, Coveny AM, Clark SM, Dunn SD. (2000) Site-directed cross-linking of b to the alpha, beta, and a subunits of the Escherichia coli ATP synthase. *J Biol Chem.* **275**, 17571-7.

28 Dunn, S. D., Cipriano, D. J., and Del Rizzo, P. A. (2004) In *Handbook of ATPases* (Futai, M. and Wada, Y., eds), pp. 311-318, Springer-Verlag, Weinheim.

29 Lupas A. (1996) Coiled coils: new structures and new functions. *Trends Biochem Sci.* **21**, 375-82.

30 Del Rizzo PA, Bi Y, Dunn SD. (2006) ATP synthase b subunit dimerization domain: a right-handed coiled coil with offset helices. *J Mol Biol.* **364**, 735-46.

31 McLachlin DT, Bestard JA, Dunn SD. (1998) The b and delta subunits of the Escherichia coli ATP synthase interact via residues in their C-terminal regions. *J Biol Chem.* **273**, 15162-8.

32 Lee LK, Stewart AG, Donohoe M, Bernal RA, Stock D. (2010) The structure of the peripheral stalk of Thermus thermophilus H+-ATPase/synthase. *Nat Struct Mol Biol.* **17**, 373-8.

33 Kumamoto CA, Simoni RD. (1986) Genetic evidence for interaction between the a and b subunits of the F0 portion of the Escherichia coli proton translocating ATPase. *J Biol Chem.* **261**, 10037-42.

34 Stalz WD, Greie JC, Deckers-Hebestreit G, Altendorf K. (2003) Direct interaction of subunits a and b of the F0 complex of Escherichia coli ATP synthase by forming an ab2 subcomplex. *J Biol Chem.* **278**, 27068-71.

35 Fillingame RH, Jiang W, Dmitriev OY. (2000) Coupling H(+) transport to rotary catalysis in F-type ATP synthases: structure and organization of the transmembrane rotary motor. *J Exp Biol.* **203**, 9-17.

36 Kumamoto CA, Simoni RD. (1987) A mutation of the c subunit of the Escherichia coli proton-translocating ATPase that suppresses the effects of a mutant b subunit. *J*

Biol Chem. **262**, 3060-4.

37 Jones PC, Hermolin J, Jiang W, Fillingame RH. (2000) Insights into the rotary catalytic mechanism of F0F1 ATP synthase from the cross-linking of subunits b and c in the Escherichia coli enzyme. *J Biol Chem.* **275**, 31340-6.

38 Dunn SD, Chandler J. (1998) Characterization of a b2delta complex from Escherichia coli ATP synthase. *J Biol Chem.* **273**, 8646-51.

39 Wilkens S, Dunn SD, Chandler J, Dahlquist FW, Capaldi RA. (1997) Solution structure of the N-terminal domain of the delta subunit of the E. coli ATPsynthase. *Nat Struct Biol.* **4**, 198-201.

40 Rodgers AJ, Wilkens S, Aggeler R, Morris MB, Howitt SM, Capaldi RA. (1997) The subunit delta-subunit b domain of the Escherichia coli F1F0 ATPase. The B subunits interact with F1 as a dimer and through the delta subunit. *J Biol Chem.* **272**, 31058-64.

41 Takeyama M, Noumi T, Maeda M, Futai M. (1988) Fo portion of Escherichia coli H+-ATPase. Carboxyl-terminal region of the b subunit is essential for assembly of functional Fo. *J Biol Chem.* **263**, 16106-12.

42 Wood KS, Dunn SD. (2007) Role of the asymmetry of the homodimeric b2 stator stalk in the interaction with the F1 sector of Escherichia coli ATP synthase. *J Biol Chem.* **282**, 31920-7.

43 Bi Y, Watts JC, Bamford PK, Briere LK, Dunn SD. (2008) Probing the functional tolerance of the b subunit of Escherichia coli ATP synthase for sequence manipulation through a chimera approach. *Biochim Biophys Acta.* **1777**, 583-91.

44 Claggett SB, Grabar TB, Dunn SD, Cain BD. (2007) Functional incorporation of chimeric b subunits into F1Fo ATP synthase. *J Bacteriol.* **189**, 5463-71.

45 Claggett SB, O'Neil Plancher M, Dunn SD, Cain BD. (2009) The b subunits in the peripheral stalk of F1F0 ATP synthase preferentially adopt an offset relationship. *J Biol Chem.* **284**, 16531-40.

46 Cherepanov DA, Mulkidjanian AY, Junge W. (1999) Transient accumulation of elastic energy in proton translocating ATP synthase. *FEBS Lett.* **449**, 1-6.

47 Panke O, Rumberg B. (1999) Kinetic modeling of rotary CF0F1-ATP synthase: storage of elastic energy during energy transduction . *Biochim Biophys Acta.* **1412**,

118-28.

48 Junge W, Pänke O, Cherepanov DA, Gumbiowski K, Müller M, Engelbrecht S. (2001) Inter-subunit rotation and elastic power transmission in F0F1-ATPase. *FEBS Lett.* **504**, 152-60.

49 Pänke O, Cherepanov DA, Gumbiowski K, Engelbrecht S, Junge W. (2001) Viscoelastic dynamics of actin filaments coupled to rotary F-ATPase: angular torque profile of the enzyme. *Biophys J.* **81**, 1220-33.

50 Wächter A, Bi Y, Dunn SD, Cain BD, Sielaff H, Wintermann F, Engelbrecht S, Junge W. (2011) Two rotary motors in F-ATP synthase are elastically coupled by a flexible rotor and a stiff stator stalk. *Proc Natl Acad Sci USA.* **108**, 3924-9.

51 Erik Kish-Trier, (2009) The peripheral stalk complex of the Thermoplasma acidophilum A1A0-ATPase: characterization, domain analysis, and interaction with the catalytic sector, Ph.D. Dissertation, Biochemistry & Molecular Biology, SUNY Upstate Medical University.

Chapter 2: Incorporation of Chimeric *b* Subunits into ATP synthase of *E. coli*

2.1 Introduction

The stator stalk of the ATP synthase in *E. coli* is a homodimer of two identical *b* subunits. The two *b* subunits form an elongated coiled coil structure which by its N-terminal hydrophobic region resides in the membrane and reaches and interacts with the top of ATP synthase (δ subunit) by its C-terminal section. On the other hand, in some bacteria and chloroplasts, the stator stalk of the ATP synthase is a heterodimer of two different *b* subunit called *b'* and *b*. These two different *b* subunits are similar to one another and also share similarity with the *E. coli b* subunit. In the absence of a high resolution structure for the stator stalk of *E. coli* a model has been proposed based on sequence and biochemical studies (1). Based on this model the two subunits make a coiled-coil structure with the two helices wrapped around each other in a right-handed manner and offset with respect to one another. This model, although unusual compared to the common left-handed coiled-coil structure, has been supported by the crystal structure of the stator stalk of a related ATP synthase (2).

As the stator stalk of *E. coli* ATP synthase is an asymmetric homodimer of identical *b* subunits, it is difficult to design experiments to study its structure, the offset model, and also interactions of the two subunits with the rest of ATP synthase; every change that is made in one of the proteins is also present in the other one and the two proteins are indistinguishable. Heterodimeric systems are advantageous for analysis of the subunit b function, but only a few ATP synthases with *b* heterodimers have been purified (3,4,5)

and the enzymes are not so well characterized, nor as easily manipulated, as that of *E. coli*. Making a heterodimeric stator stalk in *E. coli* using a chimeric approach would make it possible to differentiate between the two subunits. This would allow the design of experiments to determine the location, orientation, and specific interactions of subunits occupying the b^N and b^C positions individually through mutagenesis of each subunit independently.

As mentioned earlier Bi *et al.* (6) showed that the region between residues 55-110 in the *E. coli b* subunit can be replaced with corresponding regions of other bacterial *b* subunits without loss of function. Much of this region is in between the upper surface of F_o and the lower surface of F_1 so it is not in contact with other parts of the enzyme (7). The first attempt to make a heterodimeric stator stalk in *E. coli* ATP synthase was through a collaboration of the Dunn laboratory with Brian Cain and his student Shane Claggett. In these studies chimeric *b* subunits were generated by replacing key segments of the *E. coli uncF* gene, encoding the *b* subunit, with corresponding segments of the genes encoding either *b* or *b'* subunits of the cyanobacterium *Thermosynechococcus elongatus* (8). Sequences of two epitope tags, V5 and 6X His, was also added to these chimeric sequences to achieve specific detection. The functionality of these polypeptides was then studied by transforming the KM2 strain of *E. coli*, containing a chromosomal deletion of *uncF*, with a pair of compatible plasmids, each encoding one of the heterodimeric *b* chimeras. Some combinations of *b* and *b'* chimeras led to better growth on succinate and better proton pumping compared to results obtained when only one of the chimeras was

present. However, individual chimeras also supported both growth and proton pumping at substantial levels. Thus, the heterodimeric system was not essential for complementation.

Based on the assumption that the lack of specific heterodimerization was related to the stability of proteins from the thermophilic *T. elongatus*, the Dunn laboratory initiated similar studies using a sequences from a mesophilic source, *Rhodobacter capsulatus*. The ATP synthase of *R. capsulatus* has been previously purified and studied functionally. Furthermore, protein chemical analysis confirmed the presence of both *b* and *b'* subunits in purified ATP synthase complex (3). As in the previous approaches, the two plasmid system was used in order to co-express the *b* and *b'* subunits in *E. coli*.

In the two plasmid system each plasmid needs a separate antibiotic resistance gene and also a different origin of replication is necessary. On the other hand since each one of the introduced genes are under the control of a different promoter usually different levels of proteins are achieved which may increase the risk of homodimer formation in the cell. The other problem is the limitation of the approach in the number of plasmids transformed. Since studying the interaction of the *b* subunits with other subunits of ATP synthase are also desirable, three or even more plasmids should be used at the same time for protein chemistry analysis. This makes the system very complex and almost impossible to use effectively.

Replacement experiments in the Dunn laboratory using two plasmid system indicated that replacing residues 34-110 in the *b* subunit of *E. coli* with *R. capsulatus b* and *b'* resulted in production of a functional ATP synthase, without possibility of

homodimer formation. Neither of the the chimeric subunits by itself could support growth on succinate. In this thesis project, using previous knowledge in the lab, chimeric sequences were designed in a way that both chimeric *b* subunits and the rest of ATP synthase subunits are encoded in an operon on one plasmid. As before, this plasmid can be used in *E. coli* strain DK8 which has a chromosomal deletion of the *unc* operon (9) encoding ATP synthase. Chimeric *b* subunits were also made detectable specifically based on the sequences of tags that have been added to the proteins. Additionally, we enginered a ribosomal re-initiation site in between the two genes to foster equal levels of expression. Results presented using this system established that heterodimerization is obligatory for functionality of the system and that equal levels of the two subunits are present in the ATP synthase stator stalk.

Following the development of the chimeric heterodimeric system in *E. coli* ATP synthase, the system was used to answer some of the questions regarding the structure and orientation of the *b* subunit. Studying the orientation of the *b* subunit in the transmembrane domain was the first attempt. As determined by Del Rizzo *et al*. (1), based on evidence arising from disulfide formation studies, *b* subunit dimerization domains have a novel parallel right-handed coiled coil structure with the helices of the two *b* subunits offset by approximately one and a half turns of an α helix. One question which is still unanswered is the structure of the two *b* subunits in the transmembrane domain. In other words, we do not know if this offset relationship is local in the dimerization domain or is a phenomenon that occurs throughout the length of the stator

stalk. The ability to generate F_1F_o ATP synthase with *R. capsulatus b/b'* heterodimeric peripheral stalk provides a way to investigate this relationship and positions of the two subunits in the peripheral stalk. TID, 3-(Trifluoromethyl)-3-(m-[125I]iodophenyl)diazirine, labeling studies of the N-terminal region of the *b* showed that a contact region in *b* may include Asn2, Thr6, and Gln10 and these residues may reside on the same side of the helix (10). Other studies by use of both cross-linking and NMR techniques have shown that substitution of residues 12-21 with cysteine forms no cross-linking (11,12) and this region is not a protein-protein contact area. The only sites where cysteine substitution led to intersubunit disulfide formation were N2C, T6C, Q10C, and F14C. The developed chimeric system provided the way of mutating these positions to cysteine independently in each desired combination for further studying the orientation of these helices in this region.

The second goal for using the chimeric heterodimeric system was in studying the role of each one of the *b* and *b'* subunits in the stiffness of the stator stalk, a study which is impossible to do without having a stator stalk with distinguishable *b* subunits. A recent collaboration between our lab and Wolfgang Junge revealed that the stator stalk is stiffer than the rotor and this stiffness is necessary for the function of the enzyme (13). Introducing more than two glycine mutations in the *E. coli b* subunit disrupted the function, which emphasized the importance of stiffness in the structure of the peripheral stalk. In that study both subunits had the introduced mutation in them and it was not possible to evaluate the effect of glycine mutations in each one of the subunits

individually. On the other hand a recent crystal structure of the peripheral stalk of a related A-type ATP synthase has been published and revealed a conformation for the stator stalk in which one of the subunits seems to be straight and the other subunit bends around the straight one with help of several glycine residues in its sequence (2). The developed chimeric system now makes it possible to introduce glycine mutations in each one of the subunits individually and independently from the other to study role of each of the subunits in the stiffness or flexibility of the stator stalk.

2.2 Material and Methods

2.2.1 Synthetic sequence.

The sequence between restriction sites PacI and BssHII in pACWU1.2 (14) was used as a template for designing the chimera. These sites were used because they are unique and are located on either side of the *b* subunit gene in the vector. DNA sequences of two chimeric *E. coli b* subunits were designed and in one of them the nucleotide sequence encoding amino acids 34-110 of *E. coli* was replaced by a sequence encoding the corresponding amino acid sequence of *R. capsulatus b'* and in the other one the sequence encoding *R. capsulatus* b. Any possible unfavorable codon was changed based on *E. coli* codon usage and also codon degeneracy was used to change different positions in the DNA sequence, resulting in unique and rare restriction sites. Similar to the previous approach (8), the sequence of epitope V5 was added to the end (C-terminal) of *b'* subunit and the sequence of the epitope 6X His-tag was added to the beginning of the *b* subunit (N-terminal). Finally, the two ORFs were linked in a way that both tags are placed in

between the two, keeping the upstream area of the first *b* subunit gene and the second b-δ inter-geneic region intact. To link the expression of the two ORFs, a weak Shine-Dalgarno sequence also was added and an overlap of one base pair was used for the connection of the first gene to the second one. On the other hand, the sequence containing *R. capsulatus b'* with the V5 tag added was placed first followed by a ribosomal re-initiation region. Following this, sequence of the 6X His tag and the second chimeric *E. coli* b containing *R. capsulatus* b was placed. The 1227bp sequence was then ordered to the GenScript® for synthesis. The synthetic plasmid from GenScript® was called pAG2. It is a rather small plasmid (4020bp, with pBR322 origin of replication and ampicillin resistance gene) in which lots of restriction sites are unique. Accordingly, this plasmid was used as an intermediate plasmid for introducing mutations. In each case the mutated piece was then transferred to a bigger plasmid.

2.2.2 Plasmid construction.

All plasmids were prepared using standard recombinant DNA procedures (15) and all generated sequences were verified by sequencing. Plasmid pACWU1.2 is a vector containing all ATP synthase genes and encoded proteins from these genes are cysteine free to prevent any interference with the introduced mutations (14). Sequencing result showed that PacI site was absent in the pACWU1.2 vector that we had in the lab so the first goal was to introduce the PacI site in the pACWU1.2 vector. The pair of primers "forPacI" (mutagenic primer, Table 2.1) and "M13Forward" was used to perform a PCR reaction, using pSD300 (16) as a template, and the amplified piece was cut and pasted

Primer code	Sequence (5' ... 3')	Purpose
ForPacI	CTCTATTTAGTTAACGTTCTGATATTGCTCTTTA**A**TT AAAAGCAACGCTTACTACGC	Correction of PacI site in pACWU1.2
Q2stop	TTAACGCCACAATTCTGGGC**TAA**GCGATCGCGTTTGT CCTG	Mutation of position Q_{10} to stop in the *b* subunit

Table 2.1: Primers used in initial construction and characterization. The mutated position is shown in bold and is underlined.

back in pSD300 using HpaI and BsrGI (the resulting plasmid was called pAG1). At the end, pAG3 was made by inserting 430-bp BsrGI/PpuMI fragment of pAG1 into pACWU1.2 which had been digested completely with BsaI, PpuMI, and BsrGI (a three part ligation: BsaI/PpuMI and BsaI/BsrGI pieces from the pACWU1.2 digestion and the BsrGI/PpuMI piece from the pAG1 digestion). pAG5 was produced by pasting the chimera (from pAG2) into the pAG3 vector using PacI and BssHII sites. This plasmid contains both chimeric b' and b with specific tags added to each one of them. To produce pAG17 (*b'* deleted chimera) after digestion of pAG2 with AsiSI, the 3' overhang arising from cleavage was removed, and the end made blunt, using T4 DNA polymerase following by ligation of the ends to complete the in-frame deletion. The same procedure was used for preparing the pAG19 (b deleted chimera) except using BsmI site. The resulting sequences in each plasmid were pasted back into pAG3 using the PacI and BssHII sites. It should be mentioned that these two constructs are in-frame deletions of most of the sequence of each polypeptide. The remaining sequence encodes a short version of the polypeptide and the other chimeric b gene will be expressed in each case. To prepare pAG32, which carried a stop mutation at position 10 of the first chimeric gene, PCR was performed using pAG2 as template. The mutation in residue number 10 (Q_{10} to stop mutation) was introduced using a mutagenic primer called "Q2stop" (Table 2.1) containing a BstXI restriction enzyme site. Primer uncH was used as the reverse primer. The PCR product then cloned into pAG2 using the BstXI/BglII sites. The final mutated sequence was again pasted into pAG3 by using PacI and BssHII sites. All

cysteine and glycine mutations were performed by using a pair of mutagenic primers (Table 2.2 and Table 2.3) and pAG2 or pAG4 as template. pAG4 is also a small plasmid (chimera in pEXT20) which has been used as an intermediate vector for performing site directed mutagenesis. The resulting mutated sequences were cloned into pAG3 by using PacI/BssHII sites.

2.2.3 Minimal medium growth determination.

Minimal medium (M9) plates were prepared as previously described (17) containing either 0.2% glucose or 0.2% succinate as the sole carbon source. All plasmids were transformed into *E. coli* strain DK8, which contains a deletion of the entire operon encoding ATP synthase, and were grown on LB plates. One colony was picked and cells were resuspended in 0.9% NaCl. Following resuspension, cells were streaked out on succinate and glucose plates. Plates were incubated at 37 °C for 72 hours. At the end of 72 hours colonies were compared to the wild type growth and scored into three levels: +++, like wild type; ++, smaller than wild type; +, very small colony formation; ---, no growth.

2.2.4 ATPase activity and proton pumping assays.

Membrane vesicles for assessment of ATPase and ATP-dependent proton pumping were prepared from *E. coli* cells strain DK8 carrying the appropriate plasmids. Cells carrying pACWU1.2-based plasmids were grown with vigorous shaking at 37 °C in L broth containing 25 mM sodium phosphate, pH 7.0, and ampicillin at 40 μg/ml and harvested when A600 reached 3.0. Inverted membrane vesicle preparations were

Primer code	Sequence (5' ... 3')	Purpose
Bprime83F	GATCATCGCCGAGACCCGTGCGGTG**TGT**CAAAAGGACCTCG	Mutating position 83 in the b'
Bprime90F	TCAAAAGGACCTCGATGCC**TGC**ACGGCGAAAGCAG	Mutating position 90 in the b'
B83R	GTCGGCCTTGGCCTGCTCTGCCGCCAGCTG**ACA**GTCGCGCTTGGCTG	Mutating position 83 in the b
B90R	CGGCGCCCTTGAGACGGCGTGCGATGGCCTCCTTCAGGTCGGCCTT**ACA**CTGCTCTGCCGCCAG	Mutating position 90 in the b
2F	AGAACGTTAACTAAATAGAGGCATTGTGCTGTG**TGT**CTTAACGCCACAATTCTGGG	Mutating position 2 in the b'
6F	AGAACGTTAACTAAATAGAGGCATTGTGCTGTGAATCTTAACGCC**TGT**ATTCTGGGCCAGG	Mutating position 6 in the b'
10F	AGAACGTTAACTAAATAGAGGCATTGTGCTGTGAATCTTAACGCCACAATTCTGGGC**TGT**GCGATCGCGTTTGTCCTG	Mutating position 10 in the b'
2R	ACTTCATGGCGAAGAGCACGAACAGGACAAACGCGATGGCCTGGCCGAGGATTGTTGCATTCAG**ACA**CATGTGGTGGTGG	Mutating position 2 in the b
6R	ACTTCATGGCGAAGAGCACGAACAGGACAAACGCGATGGCCTGGCCGAGGAT**ACA**TGCATTCAGATTCATGTGG	Mutating position 6 in the b
10R	ACTTCATGGCGAAGAGCACGAACAGGACAAACGCGATGGC**ACA**GCCGAGGATTGTTGC	Mutating position 10 in the b

Table 2.2: Primers used in cysteine mutagenesis experiments. The mutated position is shown in bold and is underlined.

Primer code	Sequence (5' ... 3')	Purpose
bprimeone	CATCGAGGTCCTTTTG**ACC**CACCGCACGGGTCTC GGCGATGATCTTTTGCG	Mutating position 83 in the b'
bprimetwo	CATCGAGGTCCTTTTG**ACCACC**CGCACGGGTCTC GGCGATGATCTTTTGCG	Mutating position 82-83 in the b'
bprimethree	CATCGAGGTCCTTTTG**ACCACCACC**ACGGGTCTC GGCGATGATCTTTTGCG	Mutating position 81-83 in the b'
bprimefour	CATCGAGGTCCTTTTG**ACCACCACCACC**GGTCTC GGCGATGATCTTTTGCG	Mutating position 80-83 in the b'
bprimefive	CATCGAGGTCCTTTTG**ACCACCACCACCACC**CTC GGCGATGATCTTTTGCG	Mutating position 79-83 in the b'
bone	TCCTGGCTAGCTACGAGCGCAAGGCACGCGAAGT GCAGGGTCAGGCCGATGAGATCGTCGCTGCAGCC AAGCGCGAC**GGT**CAGCTGGCGGCAGAGC	Mutating position 83 in the b
btwo	TCCTGGCTAGCTACGAGCGCAAGGCACGCGAAGT GCAGGGTCAGGCCGATGAGATCGTCGCTGCAGCC AAGCGC**GGTGGT**CAGCTGGCGGCAGAGC	Mutating position 82-83 in the b
bthree	TCCTGGCTAGCTACGAGCGCAAGGCACGCGAAGT GCAGGGTCAGGCCGATGAGATCGTCGCTGCAGCC AAG**GGTGGTGGT**CAGCTGGCGGCAGAGC	Mutating position 81-83 in the b
bfour	TCCTGGCTAGCTACGAGCGCAAGGCACGCGAAGT GCAGGGTCAGGCCGATGAGATCGTCGCTGCAGCC **GGTGGTGGTGGT**CAGCTGGCGGCAGAGC	Mutating position 80-83 in the b
bfive	TCCTGGCTAGCTACGAGCGCAAGGCACGCGAAGT GCAGGGTCAGGCCGATGAGATCGTCGCTGCA**GGT GGTGGTGGTGGT**CAGCTGGCGGCAGAGC	Mutating position 79-83 in the b

Table 2.3: Primers used in glycine mutagenesis experiments. The mutated position is shown in bold and is underlined.

performed as described previously, by breaking cells using French press (13,000 psi) in a solution containing 1 mM PMSF, 50 mM sodium phosphate, pH 7.5, 5 mM $MgCl_2$, and 10% glycerol (7). Large debris were removed by centrifugation in a Beckman JA-20 rotor at 8,000×g (10,000 rpm) for 10 min. The membrane fraction was collected by sedimentation in a Beckman Ti-45 rotor at 150,000×g (38,000 rpm) for 90 min. The pellet was washed twice by homogenization in 10 mM Mops-NaOH, pH 7.5, containing 250 mM sucrose, 5 mM $MgCl_2$, 10% methanol, and sedimented as before. The final pellet was resuspended in a small volume of the same buffer (50-100 µl), frozen in liquid nitrogen, and stored at -80 °C. Membrane protein concentrations were determined by the method of Lowry *et al.* (18).

Membrane ATPase content was determined using the "bound ATPase" assay conditions (20) where 6 µg of membranes were mixed with 0.3 ml of TMK buffer (50 mM Tris-HCl pH 8, 5 mM $MgCl_2$, 300 mM KCl). Following that, 0.3 ml of the reaction mixture (8 mM ATP, 4 mM $MgCl_2$, 50 mM Tris-HCl) was added and the reaction was kept for 10 min at 37 °C. The reaction was then terminated by addition of 0.3 ml of 10% SDS and the amount of phosphate released was determined (19).

ATP-dependent proton pumping was measured by the quenching of 9-amino-6-chloro-2-methoxyacridine (ACMA) using ISA Fluorolog®-3-11 spectrofluorometer as described previously (20). Based on ATPase activity assay results, membrane samples containing 0.1 units of ATP synthase (pAG3, 83 µg; pAG5, 171 µg; pAG19 and pAG17 250 µg) in 2 ml buffer (1 mM Tricine, 300 mM KCl, 2 mM $MgCl_2$, pH 8.0 with NaOH)

containing 0.3 μg/ml ACMA in a fluorescence cuvette, were incubated for 4 min at room temperature. Fluorescence emission at 490 nm was then collected for 180 seconds using 410 excitation light (slits were set at 5 nm). ATP (1 mM) was added at time 10 seconds and uncoupler carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP) was added at time 120 seconds (2 μM).

2.2.5 ATP synthase extraction and fractionation.

Membrane vesicles (50 mg) were centrifuged (60 min at 180,000×g in Beckman TLA-120.2 rotor, 71000 rpm) and then suspended in a buffer containing 20 mM MES-Tricine, pH 7.0 (the starting buffer contained 0.2 M MES and 0.2 M Tricine, pH 7.0 with NaOH), 5 mM $MgCl_2$, 100 mM KCl, 2%(w/v) $C_{12}E_8$ (Octaethylene glycol monododecyl ether, nonionic surfactant), and 5% glycerol. After a 10-min incubation on ice, the suspension was centrifuged at 180,000×g for 1 hour (Beckman TLA-120.2 rotor, 71000 rpm). The supernatant (about 900 μl) was applied to a 9-30% glycerol gradient (30, 27, ..., 9% glycerol layers, 1.3 ml each layer), in 20 mM MES-Tricine, pH 7.0, 5 mM $MgCl_2$, 0.5 mM dithiothreitol, 0.1%(w/v) $C_{12}E_8$, and 0.02% phosphatidylcholine. After centrifugation at 200,000×g (Beckman SW 41 Ti rotor, 40000 rpm) for 5 hours, samples were collected from the bottom of gradients (15 drops for each sample).

2.2.6 Western blot.

SDS-PAGE was carried out using 15% separating gels. Western blotting was performed using polyvinylidene difluoride membranes as described previously (21). Transferred PVDF membranes then incubated with either of monoclonal anti-*b* (10-1B1,

b A_{128}-R_{138}) or anti-α (α-II) (22) antibodies or both for 1 hour at room temperature. Monoclonal antibody 10-1B1 against *b* was kindly provided by Drs. Gabriele Deckers-Hebestreit and Karlheinz Altendorf of Universität Osnabrück. Alkaline phosphatase-conjugated secondary antibody against mouse IgG (from Jackson ImmunoResearch) was used for 1 hour incubation . Substrate buffer containing 80 µl BCIP (5-bromo-4-chloro-3-indolyl-phosphate), 60 µl NBT (nitro blue tetrazolium), 100 mM Tris-HCl, pH 9.5, 100 mM NaCl, and 5 mM $MgCl_2$ was used for color development (Synaptic Systems, http://www.sysy.com/protocols/blot.php).

2.2.7 Disulfide formation.

Membranes were prepared in the presence of 1 mM DTT from *E. coli* strains DK8 carrying plasmids containing all combination of mutations in positions 83 and 90. For the second chimeric gene, b (with the 6X His-tag at the N-terminal), positions 83 and 90 were determined regardless of the residues of the leading 6X His tag (corresponding to the *E. coli* numbering). Membranes were treated with neutralized, 40 mM TCEP for 3 hours protected from oxygen. To achieve protection from oxygen a pressurized stream of nitrogen was directed to each tube for a short period of time before sealing with paraffin film. After 3 h protein samples were diluted in the disulfide reaction buffer to make concentration of 0.5 mg/ml of each membrane preparation. Disulfide reaction buffer contained 1 mM EDTA, 20 mM Tris-HCl, pH 7.5, 10% glycerol, and 100 mM NaCl. For zero time samples, two equal volumes were taken from each tube and diluted in 3X SDS sample buffer containing either 150 mM DTT or 40 mM NEM. After boiling for 5

minutes, samples were kept at -20 °C. Crosslink formation was started (for the remaining solution in each tube) by adding Cu(II) and phenanthroline to final concentrations of 0.1 mM and 1 mM respectively. At times 30 s, 1 min, and 10 min samples were taken and quenched with 3X SDS sample buffer containing 40 mM NEM, boiled, and kept at -20 °C. Another 10 minute sample was reduced by addition of 3X SDS sample buffer containing 150 mM DTT and boiling.

2.3 Results

2.3.1 Plasmid construction.

The main goal of this project was to develop a system in which the two chimeric subunits are equally expressed at relatively normal levels and assemble into ATP synthase exclusively as heterodimers. The desired situation was to have this system without any disruptive effects on expression of other subunits in the operon. For this reason we decided to have initiation of translation at the first gene and termination at the second gene as close to normal as possible. Tags were therefore placed at end of the first gene and start of the second gene, and a ribosomal re-initiation site placed between them to promote equal levels of translation. Plasmid pACWU1.2 was selected as the main vector. It is a plasmid which contains all of the *E. coli* ATP synthase subunit genes, encoding the subunits in the order of a, c, b, δ, γ, α, β, and ε and they are all cysteine free. Restriction sites in this plasmid were checked and PacI site (at the end of the c gene) and BssHII sites (at the beginning of the δ gene) were found to be unique and in the desired positions. Sequencing results of the pACWU1.2 on the other hand showed that the PacI site was

absent in the plasmid due to a A to T point mutation. The first goal was then to correct this mutation to insert the PacI site back in the plasmid. Using a pair of primers, "forPacI" (Table 2.1) and "M13F" for PCR on pSD300, the fragment was cut with HpaI and BsrGI and pasted back in pSD300 and the resulting plasmid was called pAG1. Following that, the BsrGI/PpuMI fragment from pAG1 was pasted in pACWU1.2 and pAG3 plasmid was made. pAG3 was then used to design the chimera and other following constructions.

The segment between PacI and BssHII sites in pAG3, containing the *E. coli* b gene, was used to design the chimera. As the start codon for the wild type b gene is GTG, it was decided to use this codon for designing the chimera. In each one of the chimeric subunits residues 34-110 were replaced with corresponding *b* and *b'* sequences from *R. capsulatus*. By using codon degeneracy, DNA sequences were changed in order to produce unique restriction sites in the genes encoding b and b', since having lots of unique restriction sites is handy in manipulating the DNA sequences. Some of the *R. capsulatus* codons were also changed based on the codon usage in *E. coli* to prevent any unfavorable codons in the chimera (23). Sequences of two tags were also introduced to the genes, V5 tag to the end of the first gene (C-teriminal of *b'*) and 6X His tag to the beginning of the second one (N-terminal of *b*) (8). As a result the initiation region for the first gene and *b-δ* intergenic region after the second gene are exactly like wild type (Figure 2.1, A, Figure 2.2). As it is optimal for the amounts of the two *b* subunits to be in 1:1 ratio, it was decided to put a ribosomal re-initiation sequence in between the two

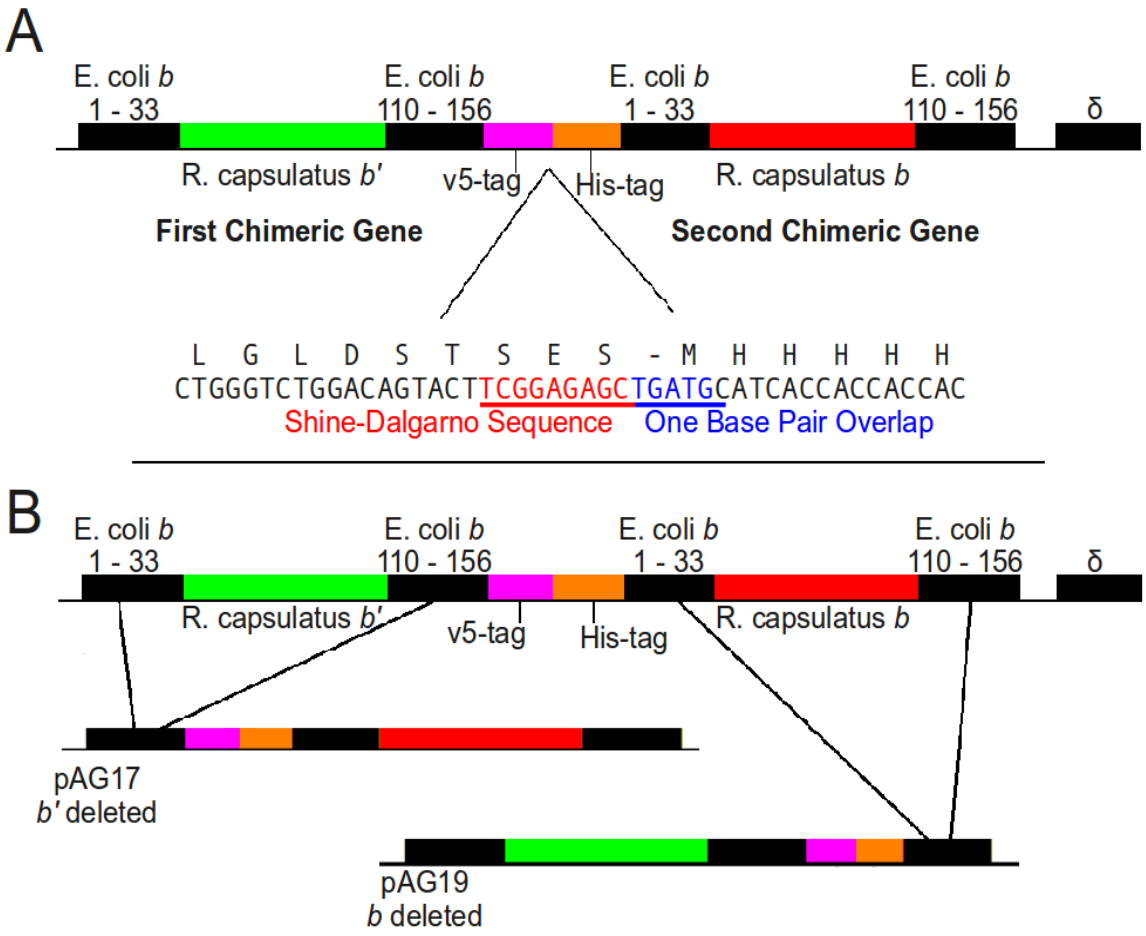

Figure 2.1: Schematic representation of the chimera and deletion constructs. *Panel A,* in the chimera residues number 34-110 of *E. coli b* was replaced with corresponding residues of *R. capsulatus b* and *b'* sequences. V5 tag was added to the C-terminal region of *b'* and 6X His tag to the N-terminal part of the *b*. To achieve equal expression for the two genes an engineered ribosomal re-initiation site was used for the region in between the two genes. Three amino acids S, E, and S were added to the end of V5 tag sequence to make a weak Shine-Dalgarno sequence. In between the two genes there is an overlap of one base pair for stop and start codon (TGATG). *Panel B,* to check the possibility of homodimer formation two deletion constructs were made. In each one of the deletion constructs the majority of the gene was deleted in a way that the rest of it was left in-frame. In each case the product of the deleted gene is a short protein which can not assemble to the ATP synthase complex.

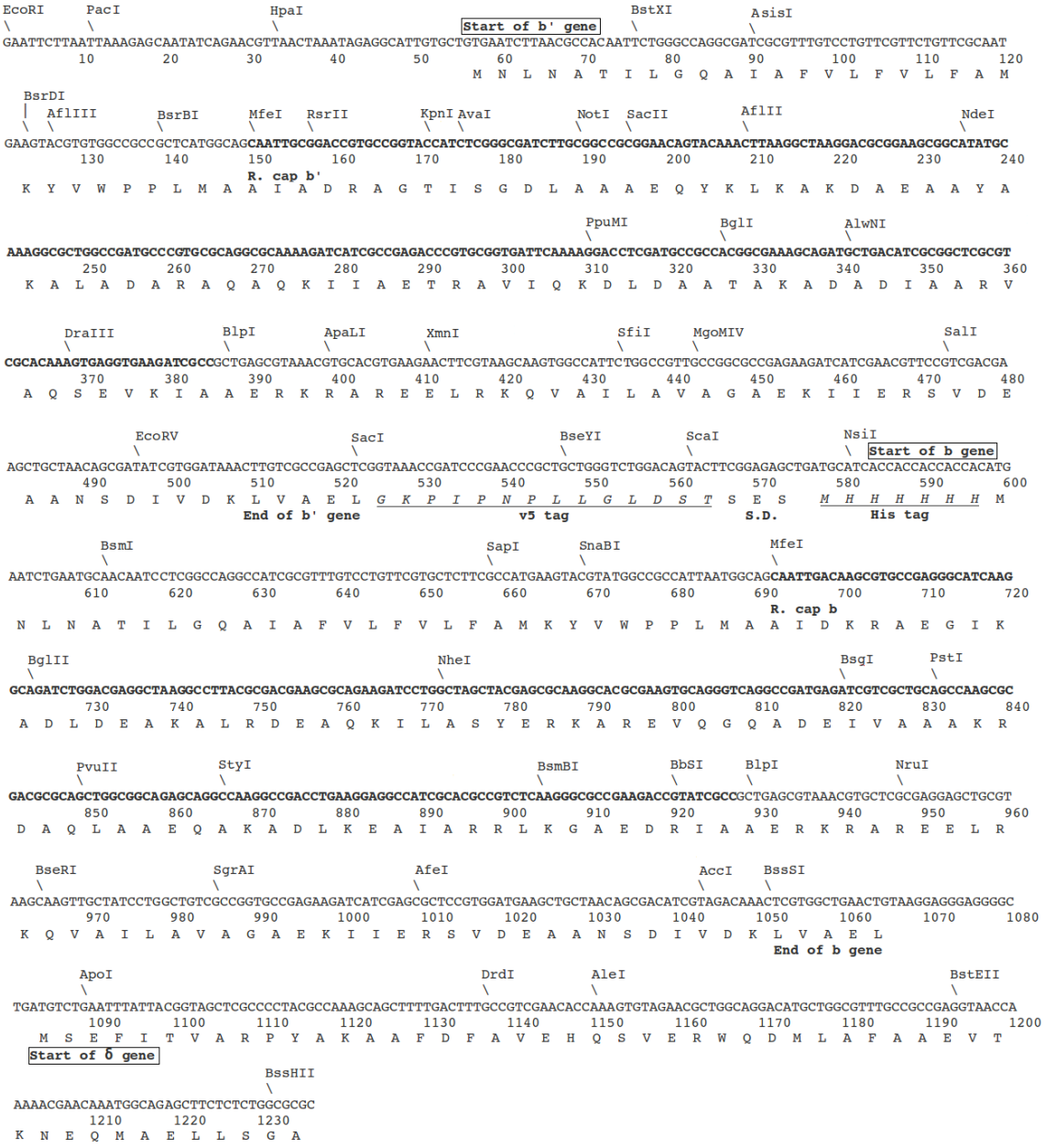

Figure 2.2: Sequence of the chimera. The sequence bounded by PacI and BssHII sites was used to design the chimera. With the help of codon degeneracy lots of unique restriction sites were introduced. Also, *R. capsulatus* codons that were bad codons for *E. coli* were changed. Sequences encoding two specific tags, V5 and 6X His were added to the b' and b respectively, for detection.

genes. The idea behind ribosomal re-initiation site is to let the ribosome stay there at its position right after finishing translation of the first gene and then let it find the start codon for the second gene and continue translating the second gene. It is used by bacteria in places where having equal amount of proteins is important (24). The criteria for having ribosomal re-initiation is to have ATG of the second gene close to the stop codon of the first gene, in some cases there is also an overlap between the end and start codon, and having a Shine-Dalgarno sequence before start codon of the next gene (25). To make a Shine-Dalgarno sequence codons encoding three amino acids were added to the end of V5 tag in the first gene, TCG GAG AGC (Figure 2.1, A). Using a strong Shine-Dalgarno sequence will increase the chance of *de novo* translation start which is not desired. Also, the stop codon of the first gene and the start codon of the second overlapped by one base in the sequence TGATG. The sequence of this designed chimera was ordered from GenScript® (Figure 2.2) and was received in a plasmid called pAG2. As mentioned earlier PacI and BssHII sites were used to clone the chimeric sequences into pAG3 producing plasmid pAG5.

To be able to test the ability of the chimeric subunits in making functional homodimers, two constructs were made in which the majority of either b' (pAG17) or b (pAG19) encoding genes were deleted leaving the rest of the gene in-frame but small in length to prevent the shortened product from assembling in the enzyme (Figure 2.1, B). All of these constructs had GTG at the beginning of the first chimera like the wild type case in *E. coli* b gene. Also, in order to check the ability of ribosomal re-initiation site to

start *de novo* translation, a stop mutation was introduced at position 10 near the beginning of the first gene, b'. This construct was called pAG32.

2.3.2 *In vivo* functionality.

As the first test for *in vivo* functionality of ATP synthase the ability of *E. coli* cells carrying the desired plasmid, pAG5, to grow on succinate was tested and compared with the controls described above. Since succinate is a non-fermentable carbon source only cells with functional ATP synthase can survive using this energy source for generating ATP. *E. coli* cells strain DK8 were transformed with the constructs to check the ability to support growth. Only chimeric b/b' expressing construct (pAG5) could grow on succinate similar to wild type. Neither *b* (pAG17) nor *b'* (pAG19) expressing constructs could grow (Table 2.4) which shows that chimeric subunits can not make functional homodimers of b/b or b'/b' form. The only functional combination is the heterodimer of b/b'. Also, construct with stop mutation at the beginning of *b'* (pAG32, Q_{10} to stop mutation) could not support growth on succinate in spite of very low expression of the b chimeric form.

2.3.3 Membrane preparation and characterization.

To investigate the effect of chimera incorporation on the function of F_1F_o and also to characterize its molecular nature, membrane vesicles were prepared for assessment of ATPase activity and ATP-dependent proton pumping. *E. coli* cells strain DK8 carrying appropriate plasmids were grown in LB media and membranes prepared as described in Material and Methods.

Detection of the two chimeric b subunits with Western blotting. In order to detect

Name of plasmid	Protein encoded	Growth on succinate	Growth on glucose
pAG3	Wild type *E coli b*	+ + +	+ + +
pAG5	Chimera of *b* and *b'*	+ + +	+ + +
pAG19	Chimeric *b'*	- - -	+ + +
pAG17	Chimeric *b*	- - -	+ + +
pAG32	Chimeric *b* (stop mutation at the beginning of b')	- - -	+ + +
pDC14	Wild type pACWU1.2 with the *b* subunit deleted (Neg. Cont.)	- - -	+ + +

Table 2.4: Complementation test on succinate for the chimera and deletion constructs. Growth of bacteria containing chimeric and wildtype plasmids used in this study on succinate (non-fermentable carbon source). In each case one colony was picked and cells were resuspended in 0.9% NaCl. Following resuspension, cells were streaked out from an LB agar plate on succinate and glucose plates. Plates were incubated at 37 °C for 72 hours. At the end of 72 hours growth of colonies were compared to the wild type growth and scored into three levels: + + + wild-type growth; + + colonies smaller than wild type; + very small colonies compared to wild type; - - - no growth.

expression levels of the two genes in membrane samples antibody specific detection of proteins was used. Several monoclonal and one polyclonal anti-*b* antibodies which were available in the lab were tested and 1B1 anti-*b* antibody was selected. The epitope for this antibody is bA_{128}-R_{138} which is common between the two chimeric *b/b'* subunits (Gabriele Deckers-Hebestreit and Karlheinz Altendorf, personal communications). The chimeric *b'* subunit (with V5 tag epitope added) and *b* subunit (with His tag epitope added) are similar in mass and number of residues. We found that they run differently upon electrophoresis and we were able to separate them completely by using 20% acrylamide gels. The result was two bands compared to one band in the wild type sample shown in Figure 2.3, A. In order to find out which band corresponds to *b'* and which to *b*, specific antibodies against V5 and His tags were used. It was determined that the lower band carries the V5 tag and is *b'* (Figure 2.3, B). The anti-His antibody recognized the upper band implying it was due to the *b* chimera, but was not very specific (data not shown). In the construct with the stop mutation at the beginning of the *b'* (pAG32, Q_{10} to stop mutation) no level of protein was detectable by Western blot which indicates that the translation initiation region of the second gene, encoding the *b'* chimera, is unable to initiate translation in the absence of translation of the upstream gene; the translational coupling is strong.

ATPase activity. ATPase activity of membranes was determined under conditions where F_1 remained associated with F_o in the membranes. This assay showed that cells carrying wild type plasmid, pAG3, have the highest levels of ATPase activity(2.21

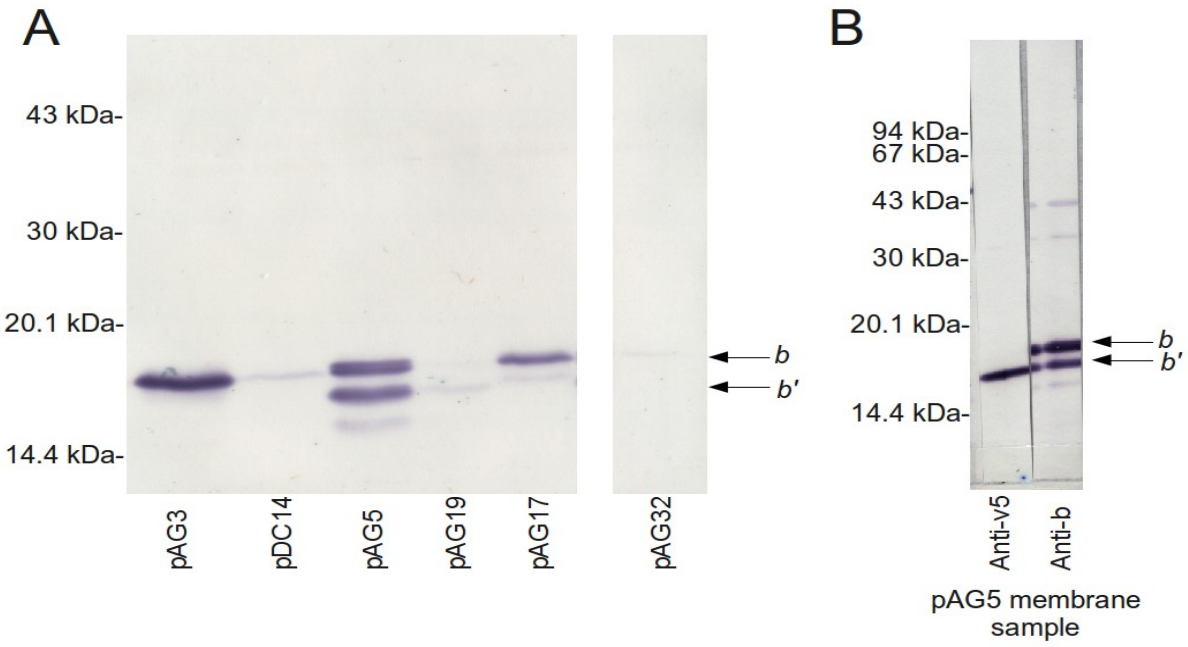

Figure 2.3: Detection of chimeric *b* subunits by 1B1 monoclonal anti-*b* antibody. Membrane proteins were subjected to Western blot analysis. The same amount of protein were loaded in each lane (5 µg). In case of chimera, pAG5, there are two bands present in the blot. The two chimeric bands were further detected by specific anti-V5 antibody. The lower band has the V5 epitope and is the *b'*. pAG3 - wt, pDC14 - *uncF* deleted, pAG5 - chimeric *b'/b*, pAG19 - *b* deleted, pAG17 - *b'* deleted.

U/mg). Compared to the wild type, the chimera, pAG5, has substantial amount of ATPase activity (1.25 U/mg) while when only one of the genes are present, pAG19 and pAG17, the activity was strongly reduced (0.16 U/mg and 0.30 U/mg respectively). As expected, the deletion control, pDC14 showed no significant ATPase activity (0.03 U/mg).

ATP-Dependent proton pumping. To investigate whether the ATPase activity of these membranes is coupled to proton translocation, the ability of ATP synthase to establish a proton gradient across the membrane was tested by measuring the ATP-dependent quenching of ACMA (9-amino-6-chloro-2-methoxyacridine) fluorescence (Figure 2.4). In normal conditions, where there is no potential gradient across the membrane, ACMA can diffuse in and out of vesicles. Addition of ATP to the solution causes the ATP synthase to work in the direction of ATP hydrolysis and since vesicles are everted, the enzyme pumps protons inside vesicles. ACMA now can go inside the membranes but it gets protonated and is trapped there. This can be monitored by monitoring fluorescent quenching following the addition of ATP. At the end, the uncoupler FCCP dissipates the proton gradient across the membrane and reverts the quenching back to the basic level. The levels of quenching were strong and similar in both cases of wild type and pAG5 chimera construct. In comparison, there was no quenching in the cases of internal deletions, despite their detectable levels of ATPase activity. This indicates that in these membranes the ATPase was not coupled.

2.3.4 ATP synthase extraction and fractionation by glycerol gradient centrifugation.

Achieving equal amounts of expression for the two subunits is an optimal situation

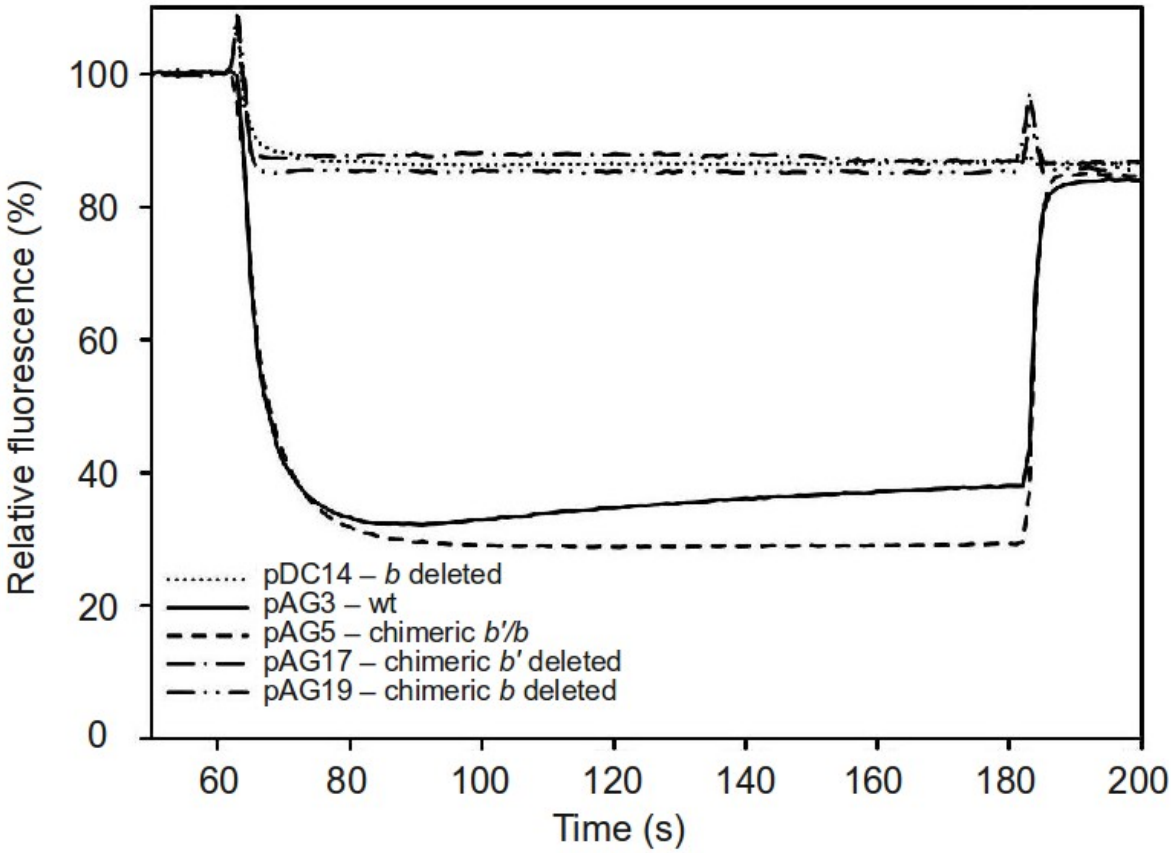

Figure 2.4: ATP-dependent proton pumping assay. Based on ATPase activity assay results, membrane samples containing equal units of ATP synthase (0.1 U) were diluted in the buffer and ACMA in ethanol added. ATP was added to the samples and the fluorescence was monitored after addition of ATP. Membranes carrying wild type and *b'/b* chimera showed almost the same amount of quenching of ACMA upon addition of ATP. Amount of membrane that used: pAG3 83 µg, pAG5 171 µg, pDC14, pAG19, and pAG17 250 µg. Finally, by addition of uncoupler FCCP the established proton gradient was dissipated to the basic level.

because it will decrease the risk of homodimer formation between the subunits. Also it will show the efficiency of cross-linking experiments when this system is used for disulfide formation studies. For checking 1:1 ratio of the two proteins we tested the expression level of the two chimeric proteins in the membrane samples by Western blot analysis. In this method intensity of the bands was used for comparison. The best way to show equal incorporation of the two *b* subunits in ATP synthase complex is to purify ATP synthase complex and detect the two subunits. To do that, membrane samples from wild type cells and cells carrying chimeric plasmid were extracted with $C_{12}E_8$ detergent and applied to glycerol gradients to perform density gradient centrifugation. Fractions were analyzed by Western blotting to determine the distribution of the *b* subunits, and the α subunit of F_1 was used as a control. Peaks of ATP synthase are present at fractions 15-17 of wild type samples (Figure 2.5, A) and 11-13 of the chimeric samples (Figure 2.5, B), as it is observable by anti-*b* and anti-α antibodies. Furthermore, the amounts of both chimeric subunits that are associated with ATP synthase expressed from pAG5 appear to be equal, implying that they are present at a 1:1 ratio.

2.3.5 Determination of the orientation of the *b* and *b'* in the chimera.

Cross-linking experiments between cysteine residues introduced in adjacent *a* and *h* positions in the region between residues 61(*a*) and 90(*h*) in *E. coli b* has been shown to be both strong and preferential (1). Also, unpublished data from our lab showed that cross-linking of cysteines introduced in positions 83 and 90 (according to *E. coli* numbering) of *R. capsulatus b'* and *b* subunits, respectively, is very efficient. To analyze

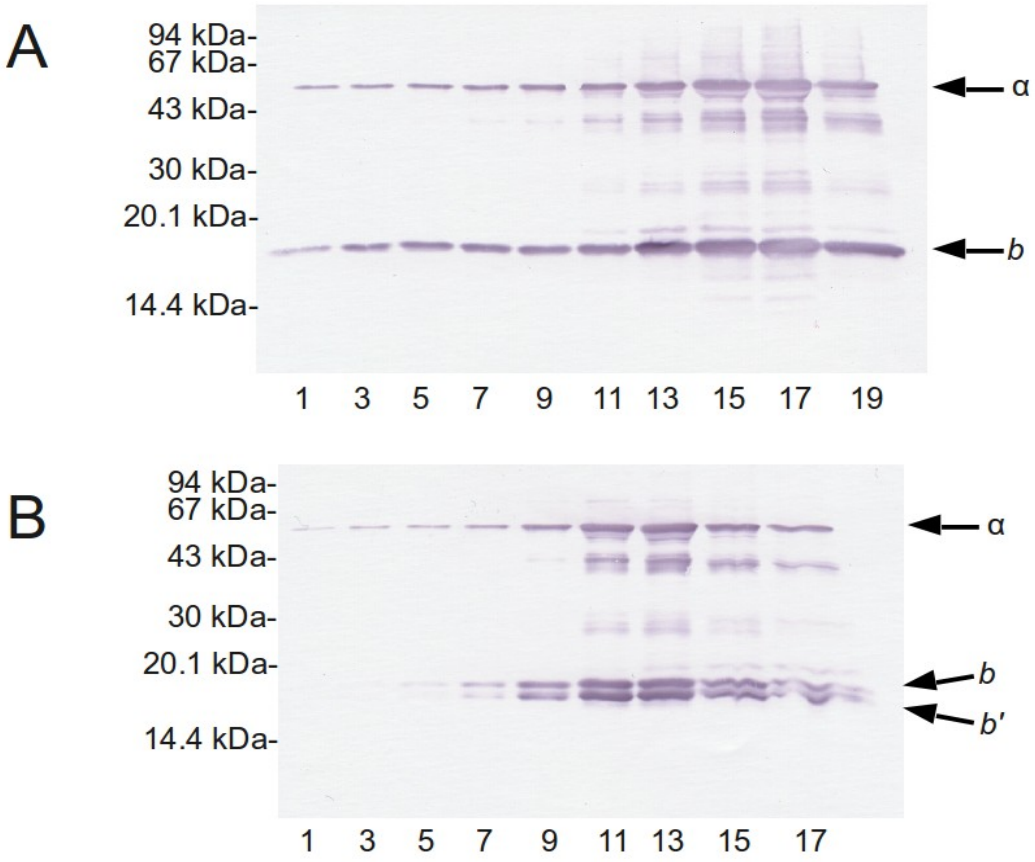

Figure 2.5: ATP synthase purification and glycerol gradient centrifugation. $C_{12}E_8$ extracted membrane proteins (7.5 mg of protein) was applied to a glycerol gradient (10-30%) containing 0.1% $C_{12}E_8$ and spun at 200,000×g for 5h. Fractions of about 400 µl were collected from the bottom (from the left, starting with number 1). The same amount of samples (15 µl) from the glycerol gradient were subjected to Western blot analysis with both anti-*b* and anti-α antibodies. *Panel A,* shows the density gradient fractionation of ATP synthase from cells carrying wild type plasmid, pAG3. Fractions 15 – 17 show the most amount of ATP synthase present. *Panel B,* fractionation of ATP synthase from cells carrying chimeric plasmid, pAG5. Fractions 11 – 13 show the highest level of ATP synthase and the intensity of the two bands are almost the same. This shows the equal incorporation of the two chimeric *b* subunits in the ATP synthase complex.

the orientation of the newly developed chimeric *b* subunits in ATP synthase of *E. coli*, the cysteine cross-linking approach was also used. Positions 83 and 90 of the chimeric subunits were mutated to cysteine in all possible combinations, b'_{83}/b_{83}, b'_{83}/b_{90}, b'_{90}/b_{83}, and b'_{90}/b_{90} to study the disulfide formation potential. The first test for these mutated constructs was the ATP synthase functionality test on succinate. As shown in Table 2.5, all combinations of cysteine mutations supported growth, although the b'_{83}/b_{83} construct grew more slowly. Secondly, disulfide formation was studied and since it requires the two cysteines to be close to each other, comparing the combination of cysteine mutations which makes the cross-link most efficiently indicates which orientation is present. Membrane samples containing the cysteine mutations were treated with Cu(II)-phenenthroline based on the method described in Materials and Methods. Samples were taken at different time intervals and were subjected to Western blot analysis with anti body 1B1 against the *b* subunits. The most efficient heterodimer formation was observed for the b'_{83}/b_{90} sample as seen by strong dimer band (Figure 2.6). In case of other combinations, despite the long incubation at 30 °C there was no significant heterodimer formation observable. The interesting observation is that the monomer bands in case of b'_{83}/b_{90} mutations disappeared completely during the time of the experiment. This results confirms that each ATP synthase complex contained one *R. capsulatus b* chimera and one *R. capsulatus b'* chimera. In all cases, DTT, a reducing agent, restored the *b* subunits to their initial levels, confirming the disulfide nature of any products.

Name of plasmid	Protein encoded	Growth on succinate	Growth on glucose
pAG3	Wild type *E coli b*	+ + +	+ + +
pAG5	Chimera of *b* and *b'*	+ + +	+ + +
pDC14	Wild type pACWU1.2 with the *b* subunit deleted (Neg. Cont.)	- - -	+ + +
pAG46	$b'_{83} - b_{83}$ cysteine mutations	+ +	+ + +
pAG47	$b'_{83} - b_{90}$ cysteine mutations	+ + +	+ + +
pAG48	$b'_{90} - b_{83}$ cysteine mutations	+ + +	+ + +
pAG49	$b'_{90} - b_{90}$ cysteine mutations	+ + +	+ + +
pAG50	$b'_{2} - b_{2}$ cysteine mutations	- - -	+ + +
pAG51	$b'_{2} - b_{6}$ cysteine mutations	+	+ + +
pAG52	$b'_{2} - b_{10}$ cysteine mutations	- - -	+ + +
pAG53	$b'_{6} - b_{2}$ cysteine mutations	- - -	+ + +
pAG54	$b'_{6} - b_{6}$ cysteine mutations	- - -	+ + +
pAG55	$b'_{6} - b_{10}$ cysteine mutations	- - -	+ + +
pAG56	$b'_{10} - b_{2}$ cysteine mutations	+ + +	+ + +
pAG57	$b'_{10} - b_{6}$ cysteine mutations	+ + +	+ + +
pAG58	$b'_{10} - b_{10}$ cysteine mutations	- - -	+ + +

Table 2.5: Complementation test on succinate for plasmids containing cysteine mutations. Growth of bacteria containing plasmid with cysteine mutations on succinate (non-fermentable carbon source). In each case one colony was picked and cells were resuspended in 0.9% NaCl. Following resuspension, cells were streaked out on succinate and glucose plates. Plates were incubated at 37 °C for 72 hours. At the end of 72 hours growth of colonies were compared to the wild type growth and scored into three levels: + + + wild-type growth; + + colonies smaller than wild type; + very small colonies compared to wild type; - - - no growth.

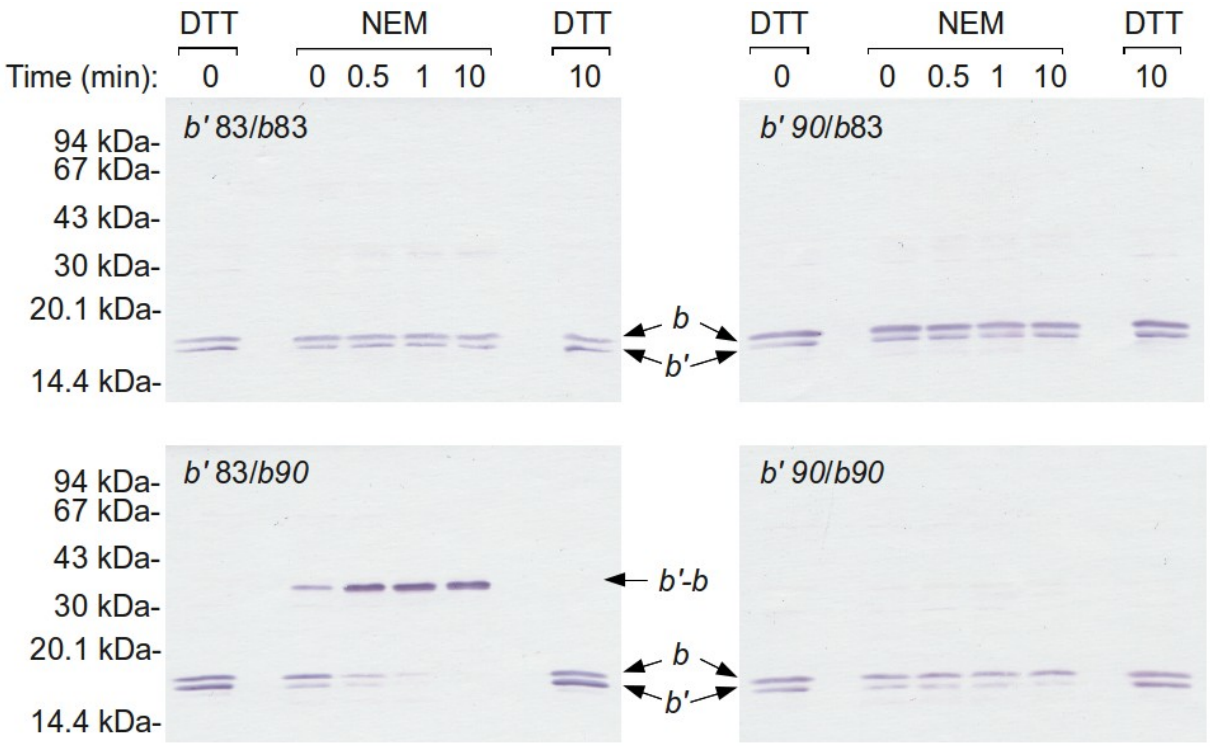

Figure 2.6: Disulfide cross-linking experiments for positions 83/90. Membranes were prepared in the presence of 1 mM DTT from *E. coli* strain DK8 carrying plasmids containing mutations in positions 83 and 90. Membranes were treated with 40 mM TCEP for 3 h protected from oxygen. Then protein samples were diluted in the disulfide reaction buffer containing 1 mM EDTA to make concentration of 0.5 mg/ml. Zero time samples were taken and diluted in 3X SDS sample buffer containing either of 150 mM DTT or 40 mM NEM. Cross-link formation was started by adding Cu(II) and phenanthroline to the final concentrations of 0.1 mM and 1 mM respectively. At times 30sec, 1min, and 10min samples were taken and quenched with NEM. At the end of 10 minutes samples were reduced by addition of DTT.

2.3.6 Studying the orientation of the *b* subunit in the transmembrane domain.

To study the offset relationship in the transmembrane domain, it was decided to start with mutations in positions 2, 6, and 10 of the chimera. Mutagenic PCR primers were designed and mutations to cysteine were introduced in all of the possible combinations of these locations: 2/2, 2/6, 2/10, 6/2, 6/6, 6/10, 10/2, 10/6 and 10/10 positions in *b'*/*b* respectively. Again the first test which is complementation test was done for all of these mutations to see their possible effect on ATP synthase functionality. To our surprise we found that majority of them did not support growth on succinate; in only two cases was growth observed (Table 2.5). Since the purpose of this experiment was to determine the orientation of each subunit in the dimerization domain it was decided to continue to the cross-linking studies despite the effect of mutations on functionality.

To do cross-linking studies membranes were made for all of these double mutants and disulfide formation was performed according to method described in Material and Methods. To detect the result of cross-linking Western blot using antibody against V5 tag was done and since only one of the chimeric *b* subunits has the V5 epitope we were expecting to see only one band in each one of the combinations. Unfortunately, not only was no significant amount of cross-linking observed, but also two bands were detected using this antibody (Figure 2.7).

2.3.7 Studying the role of the *b* and *b'* subunits in the stiffness of the stator stalk.

In this study for each one of the chimeric *b* and *b'* subunits five constructs were made in which consecutive glycine mutations were introduced in positions 79-83 while keeping the other one intact. Growth tests and studying proton pumping activities of these

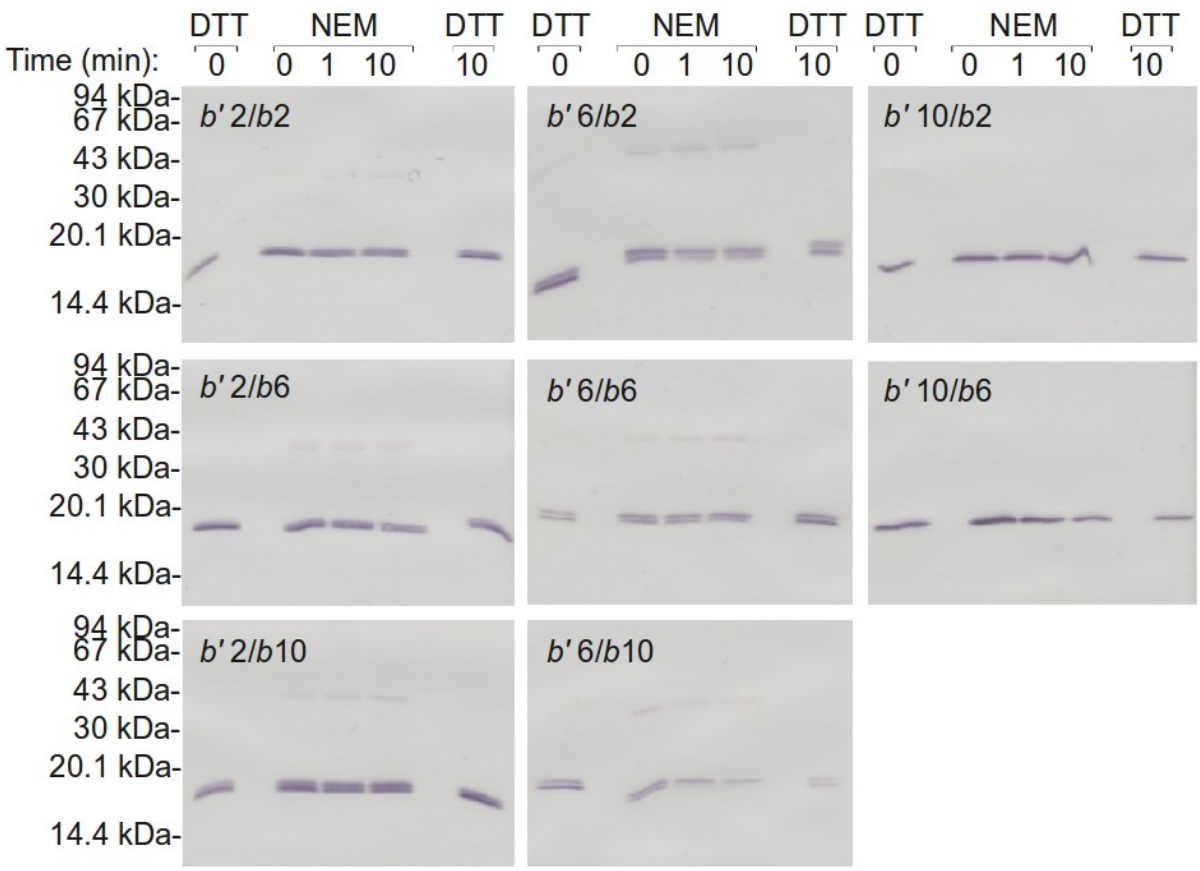

Figure 2.7: Disulfide cross-linking experiments in the transmembrane domain. Plasmids carrying all combinations of cysteine mutations in positions 2, 6, and 10 were transformed into DK8 strain of *E. coli* and membranes were prepared in the presence of 1 mM DTT. Samples of membrane were treated with 40 mM TCEP for 3 h protected from oxygen. Disulfide reaction buffer containing 1 mM EDTA was used to dilute protein samples. Concentration of 0.5 mg/ml was prepared for each sample. Zero time samples were taken and diluted in 3X SDS sample buffer containing either of 150 mM DTT or 40 mM NEM. Cross-link formation was started by adding Cu(II) and phenanthroline to the final concentrations of 0.1 mM and 1 mM respectively. At times 1min and 10min samples were taken and quenched with NEM. At the end of 10 minutes samples were reduced by addition of DTT.

constructs will help understand the role of each one of the subunits in the stator stalk stiffness. To test the functionality of the ATP synthase with stator stalk containing glycine mutations *b'* $1gly_{83}$, *b'* $2gly_{82\text{-}83}$, ... , *b'* $5gly_{79\text{-}83}$, and *b* $1gly_{83}$, *b* $2gly_{82\text{-}83}$, ... , and *b* $5gly_{79\text{-}83}$ strain DK8 of *E. coli* was transformed with each one of the constructs and the ability to support growth on succinate was tested (Table 2.6). The first observation is that introducing more than two glycine mutations into either subunit disrupts the function, *i.e.* no growth was observed. This result is similar to the effect seen with the *E. coli b* subunit homodimer, but there both subunits were affected by the mutations (13). The second observation is that bacteria with glycine mutations introduced in their *b* subunits seem to be more active than the set with the same number of glycine mutations in the *b'* subunit.

2.4 Discussion

The initial goal of this project was to produce an obligate heterodimeric system, *i.e.* an ATP synthase that will only work when the two chimeric *b* polypeptides are incorporated into the complex. The idea of making heterodimeric stator stalks in *E. coli* has been done previously in collaboration with Brian Cain's lab where *b* and *b'* genes from *Thermosynechococcus elongatus* were used to design the chimera (8). They achieved better ATPase activity and proton pumping when the two proteins were expressed but unfortunately homodimer formation was also possible. This can be explained by the fact that *T. elongatus* is a thermophile and its *b* subunits have been selected for specific heterodimer assembly and stability at elevated temperatures. However in the lab growth conditions where organisms usually are cultured at 30-37 °C,

Name of plasmid	Protein encoded	Growth on succinate	Growth on glucose
pAG3	Wild type *E coli b*	+ + + +	+ + + +
pAG5	Chimera of *b* and *b'*	+ + + +	+ + + +
pDC14	Wild type pACWU1.2 with the *b* subunit deleted (Neg. Cont.)	- - -	+ + + +
pAG59	b'_{83} glycine mutation (*b'* 1Gly)	+ +	+ + + +
pAG60	$b'_{83\text{-}82}$ glycine mutations (*b'* 2Gly)	+	+ + + +
pAG61	$b'_{83\text{-}81}$ glycine mutations (*b'* 3Gly)	- - -	+ + + +
pAG62	$b'_{83\text{-}80}$ glycine mutations (*b'* 4Gly)	- - -	+ + + +
pAG63	$b'_{83\text{-}79}$ glycine mutations (*b'* 5Gly)	- - -	+ + + +
pAG64	b_{83} glycine mutation (*b* 1Gly)	+ + +	+ + + +
pAG65	$b_{83\text{-}82}$ glycine mutations (*b* 2Gly)	+ +	+ + + +
pAG66	$b_{83\text{-}81}$ glycine mutations (*b* 3Gly)	- - -	+ + + +
pAG67	$b_{83\text{-}80}$ glycine mutations (*b* 4Gly)	- - -	+ + + +
pAG68	$b_{83\text{-}79}$ glycine mutations (*b* 5Gly)	- - -	+ + + +

Table 2.6: Complementation test on succinate for plasmids containing glycine mutations. Growth of all bacteria containing plasmid with glycine mutations on succinate (non-fermentable carbon source). In each case one colony was picked and cells were resuspended in 0.9% NaCl. Following resuspension, cells were streaked out on succinate and glucose plates. Plates were incubated at 37 °C for 72 hours. At the end of 72 hours growth of colonies were compared to the wild type growth and scored into four levels: + + + + wild-type growth; + + + colonies slightly smaller than the wild type growth; + + colonies smaller than the wild type colonies, + very small colonies compared to the wild type; - - - no growth.

homodimers might also be stable enough to form and assemble in ATP synthase complex. In this study however *b* and *b'* sequences of *Rhodobacter capsulatus* were used which is a mesophile and lives at moderate temperatures (30 °C). These might be expected to be less prone to nonspecific (*i.e.* homo-) dimerization. The other shortcoming of the previous approach was the limitation in delivery of chimeric subunits to the cells since each one of the subunits was in a separate plasmid. In contrast, in our approach, we decided to work with a single plasmid system in which the two genes were placed into *unc* operon carried on a plasmid in place of the normal *uncF* gene. And finally, to express both chimeric proteins at equal levels that are similar to that of the *E. coli* b production, we used normal contexts for initiation of the first gene, end of the second gene, and used a ribosomal reinitiation sequence to produce translation coupling between the two genes. Additionally, sequences of tags were placed in between the two genes to keep the initiation region of the first gene and the end of the second gene intact.

The result of Western blotting experiments using 1B1 monoclonal antibody against the *b* subunit detected two proteins with the expected sizes. The same level of expression for the two proteins was achieved based on the observation that the intensity of the two bands detected by Western blotting were the same. Also, we noticed that the designed ribosomal re-initiation site in between the two genes was not able to start translation *de novo*. By mutating residue number 10 in the first gene to stop codon we turned off translation of it and then we looked at the level of expression of the *b* (encoded by the second gene) by Western blot. No level of expression was detected and the construct was

not capable of supporting growth on succinate. The importance of the ribosomal re-initiation site and the linkage between expression of the first gene and the second gene can be further observed by looking at the level of expression of the *b* in case of in-frame deletion of *b'* (pAG17) compared to when the first gene is turned-off (pAG32). Since the in-frame deletion kept a small section of the b' gene in the chimera, the encoding protein (expressed by the shortened first gene) is short and can not assemble in ATP synthase but it can mimic the behaviour of a full length gene regarding the translation of these chimeric genes. Comparing the level of *b* expression in Figure 2.3, A, shows that the expression of the second gene is linked to the first gene and the region in between the two genes is a ribosomal re-initiation site instead of being a ribosomal initiation region.

Our results show that we have made an obligate heterodimeric system. Complementation tests on succinate showed that there is no chance for the chimeric subunits to form functional homodimers, only *E. coli* cells containing heterodimeric chimera were able to grow on succinate. Membrane samples from these cells showed significant amount of ATPase activity and levels of proton pumping similar to the wild type *E. coli*, something that was not achieved by using *T. elongatus* heterodimer system. On the other hand, when one of the genes was deleted, some amount of ATPase enzyme was detectable in the membrane but growth test and ATP-dependent proton pumping showed no activity. In fact, in these membrane samples the level of proton pumping was similar to the negative control.

The incorporation of the two chimeric subunits in the ATP synthase was also tested

by the enzyme extraction and fractionation. When detected by Western blotting the same intensity for the two bands was detected in fractions containing ATP synthase which shows the same amount of protein has been incorporated into the enzyme complex. To further study this, cysteine cross-linking studies were performed for positions 83 and 90 (*E. coli* numbering) of the chimera. Based on previous works in the lab we knew that in *R. capsulatus* wild type stator stalk *b'* is in b^C position and *b* is in b^N position based on the cross-linking of positions 83/90 (Mark Czuczman and Stan Dunn, personal communications). In fact, cysteine residues at position b'_{83} and b_{90} made the most efficient disulfides in the wild type *R. capsulatus b/b'*. This also provided the possibility of checking the orientation of the chimera in the *E. coli*. Our results showed that in *E. coli*, the chimera takes the same orientation as in the wild type *R. capsulatus* and we found that b'_{83}/b_{90} made the most efficient disulfide among all combinations of these positions. This is in contrast to the previous result of chimeric *T. elongatus* in *E. coli* where the most efficient disulfide was formed in b'_{90}/b_{83} construct. Also, these results further supported the 1:1 ratio of the two proteins as there is no monomer band left over after the disulfides has been formed which provides strong support for the fact that each ATP synthase molecule in the heterodimeric chimeric system has one molecule of each chimeric monomer in its stator stalk. Another interesting observation is that formation of disulfide in the newly developed system is highly specific due to the fact that the chimeric subunits were incorporated into the ATP synthase enzyme. In contrast, when the disulfide experiments are done using the expressed *b* and *b'* subunits, although b'_{83}/b_{90}

combination makes efficient cross-links, other combinations also formed to some degree (Mark Czuczman and Stan Dunn, personal communications). This confirms that the offset is indeed the preferred orientation within the ATP synthase complex and suggests that the other combinations form in isolated *b* constructs because of increased freedom or mobility in the free subunits, compared to in the complete enzyme.

As the results of this study were promising it was decided to use this system for structural studies of the stator stalk, specifically the *b* subunit. One question that we had was whether the offset model is a local phenomenon, present only in the dimerization domain or a general one also present in the transmembrane domain. Cysteines were introduced to all combinations of positions 2, 6, and 10 as mentioned earlier to test this idea. Unfortunately, many of these constructs failed to support growth on succinate, whereas Dmitriev *et al.* showed that introducing cysteine mutations in the same positions in the *E. coli* wild type homodimer did not have any effect on growth on succinate (11). Cross-linking experiments were also done but there was no significant level of disulfide formation with any combination of mutations. Also, detection of the blots with anti-V5 antibody detected two bands instead of one, since each chimeric heterodimer has one V5 epitope. The most probable explanations for this are that there could be a protein degradation process which is cleaving the chimera at a specific point preventing the protein from being functional or that the introduced cysteines are sometimes chemically modified which can also change the behaviour of proteins on gels. Another possibility might be that the 3 residues that were tested are not actually at the helix-helix interface.

Further studies would be needed to find a contact region for the two subunits in this domain since it is also possible for the two *b* subunits to not interact with each other in this region. They might be placed on either side of the *a* subunit, the membrane domains of the two subunits interacting with different faces of *a* and not with each other. Overall, because of the difficulties encountered with this approach, we decided not to pursue it further.

Another application of the system which has been started is to study the importance of stiffness or flexibility of each subunit in ATP synthase complex. Previous studies carried out in collaboration with the Junge laboratory (13) revealed that the efficiency of energy coupling by *E. coli* ATP synthase declined as the stator was made less stiff by introduction of multiple glycine residues into the region of residues 79-83. Introduction of one or two glycines was tolerated, but three glycines prevented growth on succinate, indicating that it prevented oxidative phosphorylation *in vivo*. Measurement of stator stalk stiffness by single molecule study showed that incorporation of two or three glycines produced a progressive decrease in the stiffness of the stator, revealing that a stiff stator is essential for proper enzyme function. The possibility of introducing glycine mutations in each one of the subunits independently is a great tool to study the importance of stiffness of each of the nonidentical *b* subunits. This was not possible previously because the two subunits were encoded by the same gene. Growth tests of the glycine mutation containing species showed that introduction of three consecutive glycine residues into either subunit prevented growth on succinate, just as introduction of

three glycines into both subunit of the *E. coli* homodimer had. The result shows that the stiffness of the helices of each of the subunits is important. Interestingly, however, there was better growth in the set with glycine mutations in the *b* subunit compared to mutations in the *b'*. In comparing the sequences of the two subunits, it is notable that the *R. capsulatus b* sequence has native glycines in positions 70 and 104 (*E. coli* numbering) whereas the *b'* sequence has no glycine residues so close to the point where these were introduced. Based on the results presented here, the stiffness of *b'*, which occupies the b^C position, is more critical to stator stalk function than the stiffness of *b*, which occupies the b^N position. Studying proton pumping activities of these constructs and also combining glycine mutations in the two subunits will help us understand how the combination of subunits promotes stator stalk stiffness.

2.5 References

1 Del Rizzo PA, Bi Y, Dunn SD. (2006) ATP synthase b subunit dimerization domain: a right-handed coiled coil with offset helices. *J Mol Biol.* **364**, 735-46.

2 Lee LK, Stewart AG, Donohoe M, Bernal RA, Stock D. (2010) The structure of the peripheral stalk of Thermus thermophilus H+-ATPase/synthase. *Nat Struct Mol Biol.* **17**, 373-8.

3 Gabellinia N, Gaoa Z, Eckerskorna C, Lottspeicha F, Oesterhelta D. (1988) Purification of the H+-ATPase from Rhodobacter capsulatus, identification of the F1F0 components and reconstitution of the active enzyme. *Biochim Biophys Acta - Bioenergetics*. **934**, 227-234.

4 Peng G, Bostina M, Radermacher M, Rais I, Karas M, Michel H. (2006) Biochemical and electron microscopic characterization of the F1F0 ATP synthase from the hyperthermophilic eubacterium Aquifex aeolicus. *FEBS Lett.* **580**, 5934-40.

5 Böttcher B, Schwarz L, Gräber P. (1998) Direct indication for the existence of a double stalk in CF0F1. *J Mol Biol.* **281**, 757-62.

6 Bi Y, Watts JC, Bamford PK, Briere LK, Dunn SD. (2008) Probing the functional tolerance of the b subunit of Escherichia coli ATP synthase for sequence manipulation through a chimera approach. *Biochim Biophys Acta.* **1777**, 583-91.

7 McLachlin DT, Dunn SD. (2000) Disulfide Linkage of the b and δ Subunits Does Not Affect the Function of the Escherichia coli ATP Synthase. *Biochemistry.* **39**, 3486–90.

8 Claggett SB, Grabar TB, Dunn SD, Cain BD. (2007) Functional incorporation of chimeric b subunits into F1Fo ATP synthase. *J Bacteriol.* **189**, 5463-71.

9 Klionsky DJ, Brusilow WS, Simoni RD. (1984) In vivo evidence for the role of the epsilon subunit as an inhibitor of the proton-translocating ATPase of Escherichia coli. *J Bacteriol.* **160**, 1055–1060.

10 Hoppe J, Brunner J, Jørgensen BB. (1984) Structure of the membrane-embedded F_0 part of F_1F_0 ATP synthase from Escherichia coli as inferred from labeling with 3-(Trifluoromethyl)-3-(m-[125I]iodophenyl)diazirine. *Biochemistry.* **23**, 5610-6.

11 Dmitriev O, Jones PC, Jiang W, Fillingame RH. (1999) Structure of the membrane domain of subunit b of the Escherichia coli F0F1 ATP synthase. *J Biol Chem.* **274**, 15598-604.

12 Deckers-Hebestreit G, Greie J, Stalz W, Altendorf K. (2000) The ATP synthase of Escherichia coli: structure and function of F(0) subunits. *Biochim Biophys Acta.* **1458**, 364-73.

13 Wächter A, Bi Y, Dunn SD, Cain BD, Sielaff H, Wintermann F, Engelbrecht S, Junge W. (2011) Two rotary motors in F-ATP synthase are elastically coupled by a flexible rotor and a stiff stator stalk. *Proc Natl Acad Sci USA.* **108**, 3924-9.

14 Kuo PH, Ketchum CJ, Nakamoto RK. (1998) Stability and functionality of cysteine-less F(0)F1 ATP synthase from Escherichia coli. *FEBS Lett.* **426**, 217-20.

15 Sambrook, J., Fritsh, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

16 Sielaff H, et al. (2008) Domain compliance and elastic power transmission in rotary FoF1-ATPase. Proc Natl Acad Sci USA. **105**, 17760–17765.

17 Cipriano DJ, Wood KS, Bi Y, Dunn SD. (2006) Mutations in the dimerization domain of the b subunit from the Escherichia coli ATP synthase. Deletions disrupt function but not enzyme assembly. *J Biol Chem.* **281**, 12408-13.

18 Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. (1951) Protein measurement with the folin phenol reagent. *J Biol Chem.* **193**, 265–75.

19 TAUSSKY HH, SHORR E. (1953) A microcolorimetric method for the determination of inorganic phosphorus. *J Biol Chem.* **202**, 675-85.

20 Cipriano DJ, Bi Y, Dunn SD. (2002) Genetic Fusions of Globular Proteins to the ε Subunit of theEscherichia coli ATP Synthase. *J Biol Chem.* **277**, 16782–90.

21 Stanley D. D. (1986) Effects of the modification of transfer buffer composition and the renaturation of proteins in gels on the recognition of proteins on western blots by monoclonal antibodies. *Analytical Biochemistry.* **157**, 144–53.

22 Aggeler R, Capaldi RA, Dunn S, Gogol EP. (1992) Epitope mapping of monoclonal antibodies to the Escherichia coli F1 ATPase alpha subunit in relation to activity effects and location in the enzyme complex based on cryoelectron microscopy. *Arch Biochem Biophys.* **296**, 685-90.

23 Novy R, Drott D, Yaeger K, Mierendorf R. (2001) Overcoming the codon bias of E. coli for enhanced protein expression. *InNovations.* **12**, 1–3.

24 Das A, Yanofsky C. (1984) A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. *Nucleic Acids Research.* **12**, 4757–68.

25 Eyre-Walker A. (1996) The close proximity of Escherichia coli genes: Consequences for stop codon and synonymous codon use. *Journal of Molecular Evolution.* **42**, 73–8.

Chapter 3: Bioinformatics Analysis of *b* and *b'* Sequences of ATP Synthase

3.1 Introduction

The ATP synthase in *E. coli* has a homodimeric stator stalk (as do most *Bacteria*), while many other organisms have a heterodimeric stator stalk composed of two *b* type subunits *b* and *b'*. Organisms with heterodimeric *b* subunits include chloroplasts and photosynthetic *Bacteria* such as *Cyanobacteria* and certain *Proteobacteria*. These genes often are referred to as ATPF and ATPX for *b* and *b'* respectively and sequence alignment of these proteins within an organism show significant amounts of divergence. Different criteria has been used for the purpose of assigning each one of these proteins to a sub-family. The historical convention of assigning the longer sequence as *b* and the shorter one as *b'* (1) has been used as one criterion. It should be mentioned that *Cyanobacteria* were the first organisms in which their heterodimeric *b* subunits were discovered as *b* and *b'*. Alternatively, studying the structure of the dimerization domain of the homodimeric *b* subunit in *E. coli* (by use of cysteine cross-linking experiments) has lead to the assignment of b^N and b^C subunits (2). Cysteine cross-linking experiments showed that the two subunits, instead of being in-register, are offset with respect to one another. Based on this the subunit which was shifted toward the N-terminal was called b^N and the other one was called b^C. Further studies using engineered subunits fixed in specific orientations revealed different roles for each subunit and different interactions with the ATP synthase complex (3). Overall, based on these studies we now know that the two subunits in the stator stalk have different interactions with each other and the rest of the ATP synthase

subunits. Accordingly, our expectation is that the differences between the two subunits in a heterodimer will reflect this. How these interactions are correlated with existing *b* and *b'* designations, based purely on length, is unknown. While previous studies examined the difference between the two subunits (4,5) they did not examine the evolution of these proteins nor did they inform on how to distinguish among these proteins based on sequence. In other words, it is still not known if length is an appropriate criterion for the assignment. The aim of this project was to study sequence elements responsible for *b* and *b'* nomenclature (or occupation of b^N or b^C positions) corresponding to *b/b'* length and possibly, find out how these heterodimers arose. Since we expect *b* and *b'* arose through gene duplication followed by divergence, determining the points in evolution where the heterodimer arose and the chance of separation between organisms is high will allow identification of clades with common characteristics.

At first it was necessary to find a good way of classification of these organisms based on common properties. Literature searches showed that gene arrangement of ATP synthase operon can be a good candidate for this purpose since at least three different patterns have been reported among *Bacteria*. In *E. coli* (6), the alkaliphilic bacterium *Bacillus firmus OF4* (7), and the thermophilic *Bacillus strain PS3*(8) the ATP synthase genes were reported to be grouped together forming a single transcriptional unit encoding the *a*, *c*, *b*, δ, α, γ, β, and ε subunits. However, in the purple nonsulfur photosynthetic bacteria *Rhodopseudomonas blastica* (9), *Rhodospirillum rubrum* (10), and *Rhodobacter capsulatus* (11,12) the ATP synthase operon was shown to be separated into two separate

groups; one operon encoding the F_o (*a*, *c*, *b'*, and *b*) and the other operon encoding the F_1 (δ, α, γ, β, and ε) subunits. And finally, the last reported pattern was found among *Cyanobacteria Synechococcus strain PCC 6301* (1), *Synechococcus strain 6716* (13), *Synechocystis sp strain PCC 6803* (14), and *Anabaena strain PCC 7120* (15). Here the genes again were arranged into two transcriptional units, however, this time one containing the genes encoding a, c, b', b, δ, α, and γ and the other containing genes for the β and ε subunits. Since these splittings of operons were different and were likely to be rare and unique events in these cases, they were selected as an initial approach of analyzing the organization of genes in ATP synthase transcriptional units.

The availability of a large number of completely sequenced bacteria provided a large database to start with. Several different patterns of operon arrangements were found among *Bacteria*. Among these patterns the ones that were rare, meaning a small number of bacteria had those kinds of patterns, were excluded. In contrast, patterns that were more common were selected for further studies. Subsequently, identified patterns were put onto a tree of *Bacteria*, determined by ribosomal protein sequence alignments, to see the distribution of each operon arrangement pattern among different phyla of *Bacteria*. Following that, the relationships between the two *b* subunits, defined by gene order, was determined by making phylogenetic trees of *b* subunit sequences. At this point, since there are several groups of *b* subunits available (based on operon arrangement patterns), sequence-based characteristics of any given group can be identified through development of position-specific scoring matrices (PSSM) (16) for those groups. These matrices will

be used for BLAST (17) search to find other related sequences in the database. Finally, discovered patterns will be helpful for finding sites of interaction (both inter- and intra-molecular) in the dimer of *b* subunits by coevolution analysis (18).

3.2 Material and Methods

3.2.1 Sequence search and graphing.

All completely sequenced organisms by the end of August 2011 were downloaded as Genbank files and searched for the ATP synthase genes using BioPython (19) programming language and PSI-BLAST (20) search. The SeqIO library of the BioPython language provides parsing through genbank files and looking for specific sections in the file. For example in this study "product" in the CDS (coding sequence) under the FEATURES section in each organism's genbank file was searched to find F-type ATP synthase genes. Following this step, to make sure that the found sequences are correct, their similarity with the *b* subunit of *E. coli* was confirmed by running a PSI-BLAST search by use of the *b* subunit of *E. coli* as a query and collecting the found hits. PSI-BLAST was used to be able to find all *b* subunits, even distantly related ones. Search was continued until convergence, usually less than 10 iterations with E-value of 10^{-10}. Comparing the results of the two methods ("product" name search and PSI-BLAST search) helped to make sure that the found sequences are really related to the *b* subunit (both by annotation and sequence similarity). Graphical figures were made using the GenomeDiagram (21) library of the BioPython language. This library adds the name of every found sequence to a list of hits in each organism's genbank file and at the end

graphs all found hits in one page. It prints all of the genome in one page (pdf file or an image file), with the found hits high-lighted as coloured arrows or boxes. The colour code used in this study to relate genes with their encoded polypeptide is: a, red; c, blue; b or b', green; delta, light-blue; alpha, orange; gamma, purple; beta, dark-sea-green; and epsilon, yellow.

3.2.2 Phylogenetic analysis.

Multiple sequence alignments were constructed using the MUSCLE (22) program. In each case a collection of sequences were selected and aligned with this program with the default settings. After the multiple sequence alignment, their curation was done by Gblocks (23). Gblocks eliminates poorly aligned positions and divergent regions of an alignment of DNA or protein sequences prior to phylogenetic analysis. The default options of Gblocks were used. Maximum-likelihood phylogenetic analyses were performed using both PHYML (24) and RAxML (25). For PHYML the default setting was used. With RaxML 100 steps of bootstrapping followed by 20 steps of maximum-likelihood search was performed. The output of the program is the best tree that RAxML can make using the 20 maximum-likelihood cycles. Tree rendering was performed by FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

3.3 Results

3.3.1 Searching for ATP synthase genes.

The whole genome data for completely sequenced organisms (1441 *Bacteria*), was downloaded from NCBI Genome database

(http://www.ncbi.nlm.nih.gov/books/NBK3837/) by August, 2011 and the BioPython (19) library for the Python programming language was used to parse through the genebank files and search for desired names. *Archaea* and some organisms among *Bacteria* have both V-type and F-type ATP synthases and some of them only have V-type ATP synthase. Since the focus of this study was on F-type ATP synthase, all *Archaea* and *Bacteria* with V-type ATP synthase were excluded using product name search (Figure 3.1). There are numerous errors in annotations of newly sequenced genes and the programs used only searched for product names in genbank files and therefore there is a chance of misclassification in some of the steps. To lower the risk of errors a PSI-BLAST search was also performed to confirm the result of the name searches. This step also helped to find organisms where the *b*-δ genes were fused into a single gene. Since there is only one gene called "*b* subunit" in these organisms thus, a name-based search can find one subunit, but a BLAST search can find two hits in which the second one is *b*-δ fusion. The graphical representations of *unc* operons (all genes encoding ATP synthase subunits) were made in each bacterium using GenomeDiagram (21) (Figure 3.2) and examined one by one to find the gene structure for each organism. Both position and order of genes were studied in each case. It should be emphasized that in nearly all cases the order of ATP synthase genes is the same and there is no difference among organisms. However, in some species of *Streptococcus* (23 organisms) gene c precedes gene a. On the other hand, based on the split in the ATP synthase operon explained in the introduction, organisms are very diverse and can be classified into several different categories.

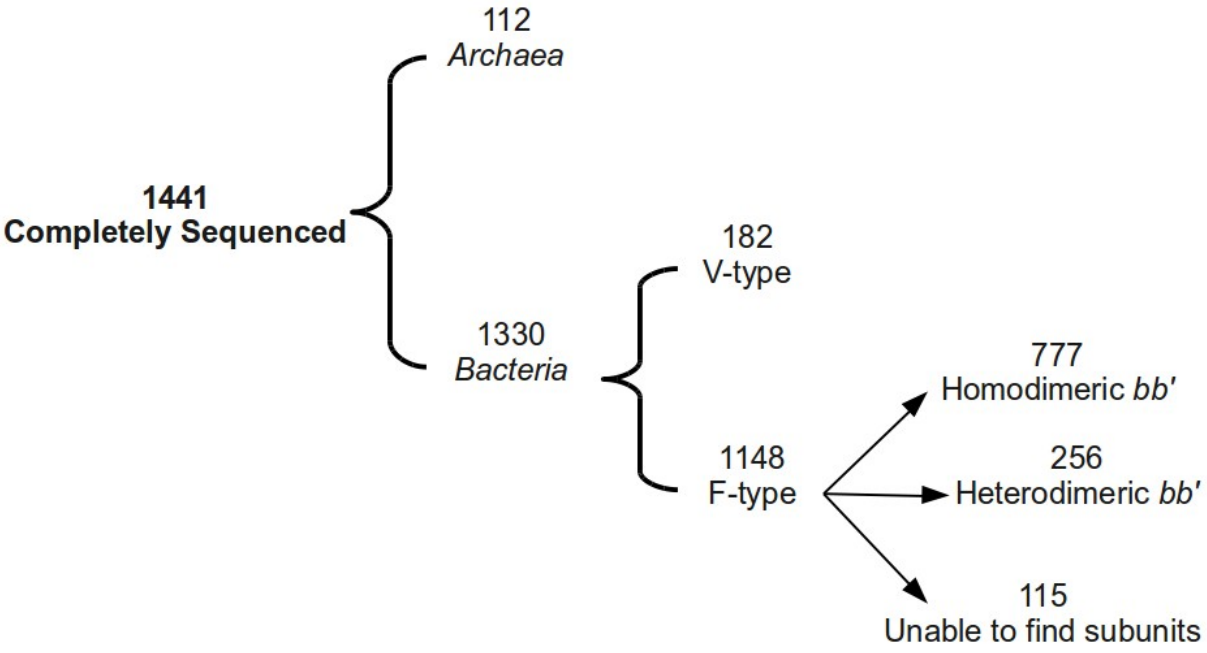

Figure 3.1: A summary of the completely sequenced organisms used in this study. By the end of August, 2011, all completely sequenced organisms were downloaded from the NCBI database. Among 1441 downloaded organisms, 1330 *Bacteria* were found (*Archaea* were excluded). 1148 *Bacteria* which had F-type ATP synthase were selected to work with. Among these 777 organisms had a homodimeric stator stalk while the stator stalk of 256 other organisms was heterodimeric. In 115 cases because of lack of information or annotation errors it was not possible to find sequences.

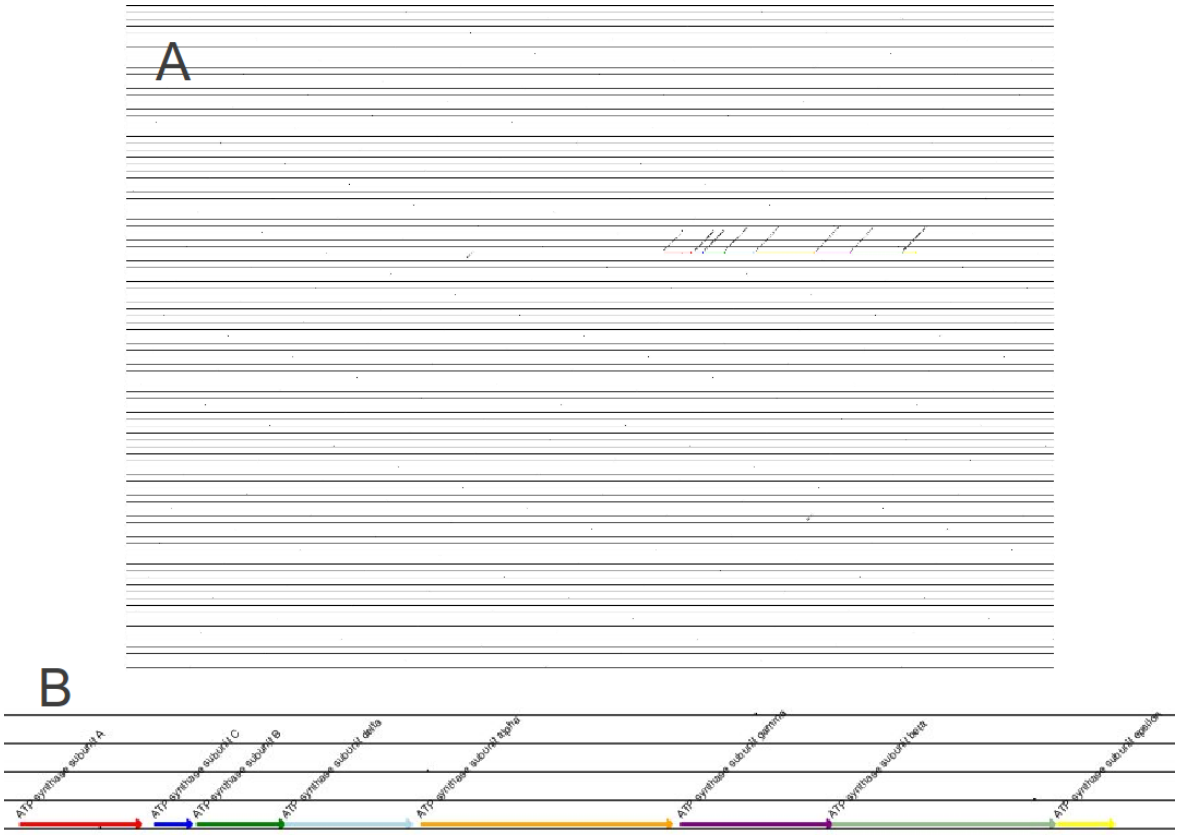

Figure 3.2: The output of the program used for graphing ATP synthase genes. The BioPython GenomeDiagram library prints all genome of an organisms in one page, with the interested genes highlighted as an arrow or box (*Panel A*). It is also possible to add different colours to distinguish among genes more easily. The direction of arrows show the way transcription proceeds (*Panel B*). The colour code used in this study: gene a is red; gene c is blue; genes b and b' are green; gene delta is light-blue; gene alpha is orange; gene gamma is purple; gene beta is dark-sea-green; gene epsilon is yellow.

3.3.2 ATP synthase gene structure.

For the majority of homodimeric organisms all of the ATP synthase genes are next to each other in one complete operon (Figure 3.3, A). In contrast, in the case of heterodimers usually there is a split in the operon which divides it into two (Figure 3.3, B) or three (Figure 3.3, C) separate units. Based on these patterns, genomes with *b* subunit heterodimers were divided into 5 different groups which has been shown in the Figure 3.4. An interesting observation is in the phylum of *Actinobacteria*, which is the only phylum that has both homo and heterodimeric *b/b'* organisms. In this phylum, two organisms with heterodimeric *b* subunits were found to have all genes in one single operon: *Segniliparus rotundus* and *Pseudonocardia dioxanivorans*. Other cases were found where a fusion exists between the second *b* subunit gene and the δ gene, so they also make one complete single operon: *Mycobacterium* species, *Gordonia bronchialis*, and *Tsukamurella paurometabola.*

In this study we worked on the 5 major patterns that were more common among organisms (Figure 3.4). There are also other types of patterns which are not very common among organisms and only a few of them have those types in their ATP synthase gene structure (Figure 3.5). These were excluded from the study.

3.3.3 Phylogenetic tree of *Bacteria*.

To study the evolution of these patterns among organisms it was decided to look at how these species map onto a phylogenetic tree of *Bacteria*. Our assumption was that gene rearrangements are rare and stably maintained once they arise. Whether or not these patterns are the result of an independent gene duplication event will further help to

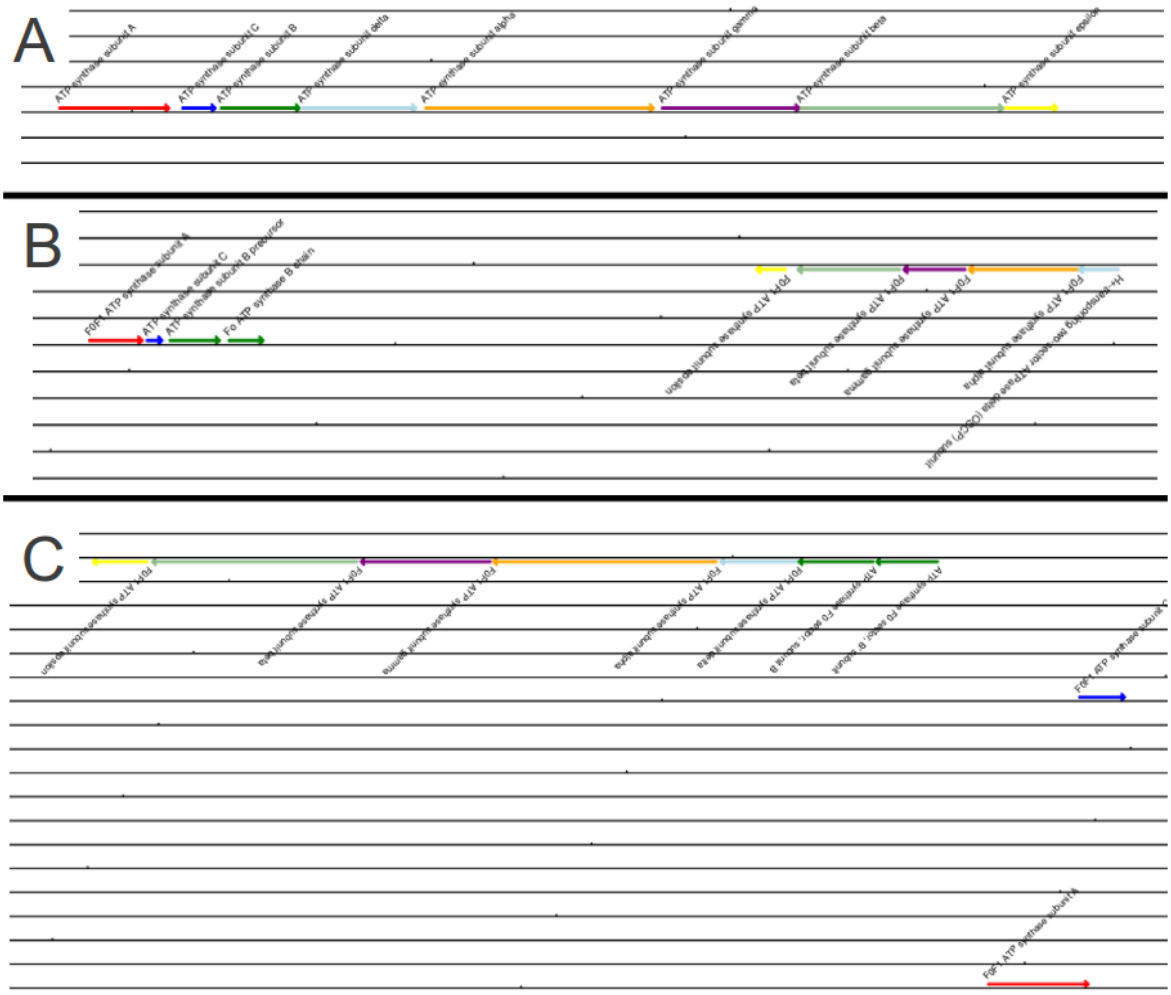

Figure 3.3: Representations of different kinds of ATP synthase gene structure. *Panel A,* shows the pattern found in the majority of homodimers. In this gene structure there is only one operon encoding the ATP synthase subunits and all of the genes are located next to each other without any split in between them. In contrast heterodimers usually have a split in their operon which divides the operon into several transcription units; two in *Panel B* and three in *Panel C*.

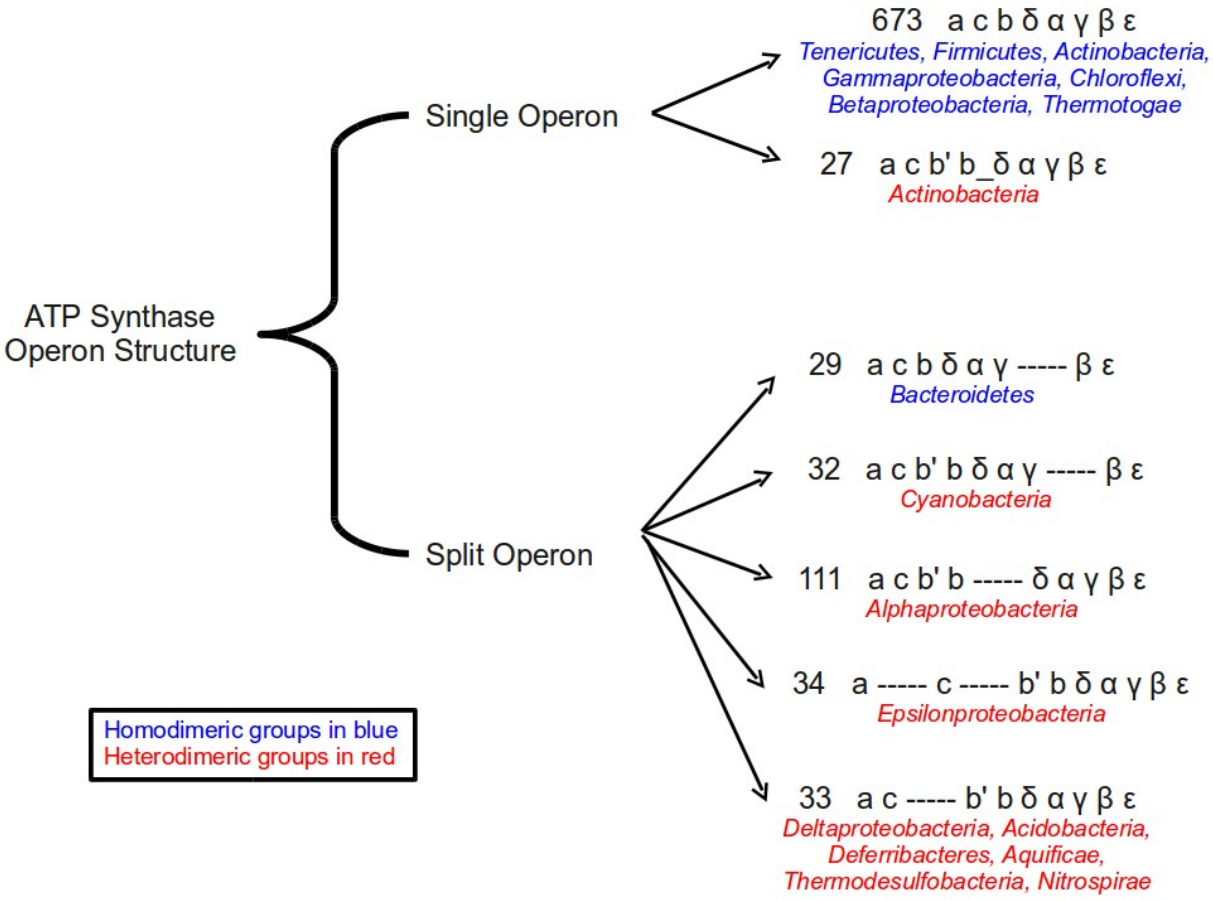

Figure 3.4: A summary of different patterns found in this study. This figure shows the classification of *Bacteria* based on the ATP synthase operon patterns and the number of organisms in each group. In general, organisms can be divided into single and split operon patterns which in the latter case the operon is divided into two or more different transcriptional units. Organisms with homodimers usually have a single operon with all of the genes next to each other making one transcriptional unit. The exception to this is the phylum of bacteroidetes which has the same structure as *Cyanobacteria* (heterodimeric), with beta and epsilon genes making one separated unit. Organisms with heterodimers on the other hand, usually have a split (shown by "-----") in their ATP synthase operon. The exception to this is the phylum of *Actinobacteria* (shown in both red and blue) which is the only phylum in which both homodimeric and heterodimeric bacteria are present. The heterodimeric ones in this phylum usually have a fusion between the b and delta genes (shown by underscore "_") while the b' gene is separated with a different open reading frame. Number of bacteria in each group is shown in each pattern.

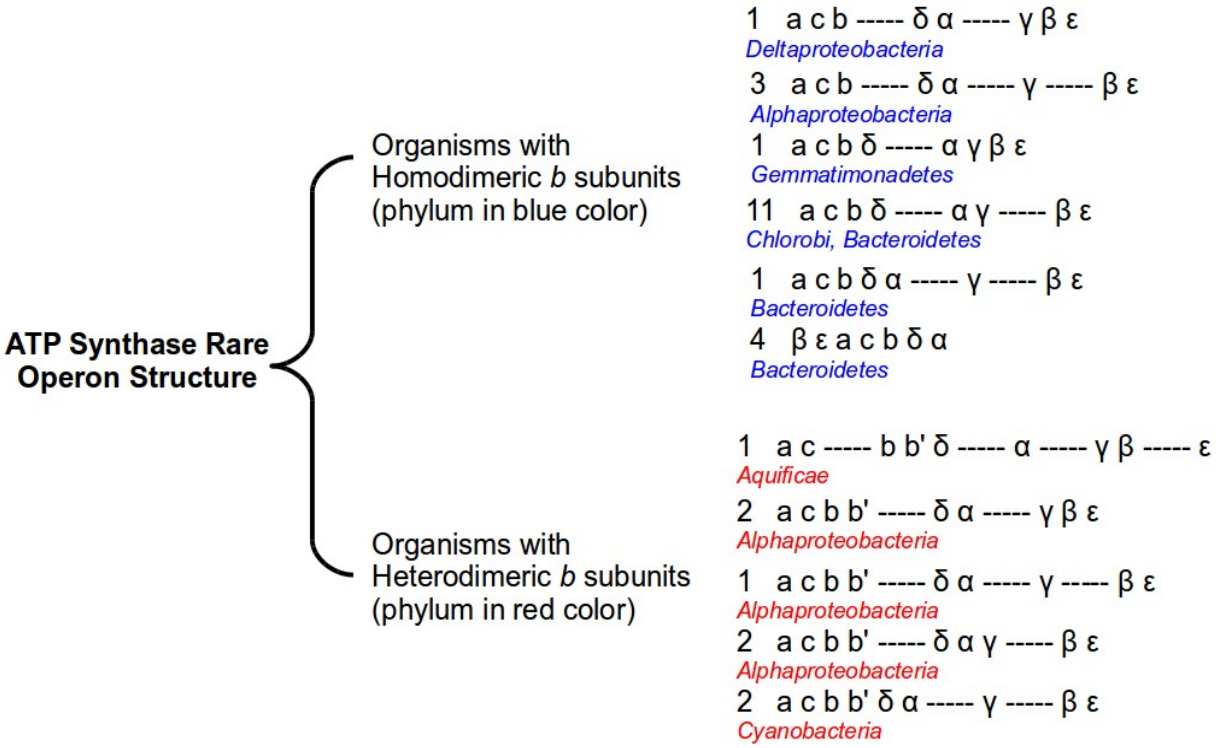

Figure 3.5: A summary of rare patterns found in this study. A summary of rare patterns in this study is presented in this figure. The number of bacteria in each group is shown. The split in the ATP synthase operon is shown by "-----". These bacteria were excluded from this study since their number is so low.

classify these organisms and study their possible similarities. To make the tree of *Bacteria*, ribosomal proteins were selected since these proteins were very ancient and well conserved among organisms and they tend not to be horizontally transferred (26). Both large and small ribosomal subunits can be used and there is no preference for large over small ribosomal proteins. In this study proteins of large ribosomal subunits were used. The tree was rooted by including five organisms from archaea (2 from Crenarchaeota and 3 from Euryarchaeota). Having outliers in the database helps to pinpoint the position of the root on the tree. To be able to perform the phylogeny, large subunits of ribosomal proteins that are common among both *Bacteria* and *Archaea* was collected for each organism (27). For each organism in the study protein sequences of 12 large ribosomal subunits (L1-L5, L10-L11, L13-L15, L18, and L22) were selected (by both annotation search and PSI-BLAST search) and concatenated. These sequences were used to make multiple sequence alignments by MUSCLE (default setting). To be able to use informative residues in the alignments, they were further curated with Gblocks (23) (columns with conserved sequences were kept, poorly aligned columns were deleted). Then a bootstrapped maximum-likelihood tree was made using RAxML and was coloured according to operon structures. Figure 3.6 shows the tree of *Bacteria* with heterodimeric organisms highlighted. Bacteria carrying each of the distinctive rearrangements of the ATP synthase genes into multiple transcription units cluster together, indicating their evolutionary relationship. The distribution of these clades imply that they likely arose from multiple gene duplication events. Assuming that the initial

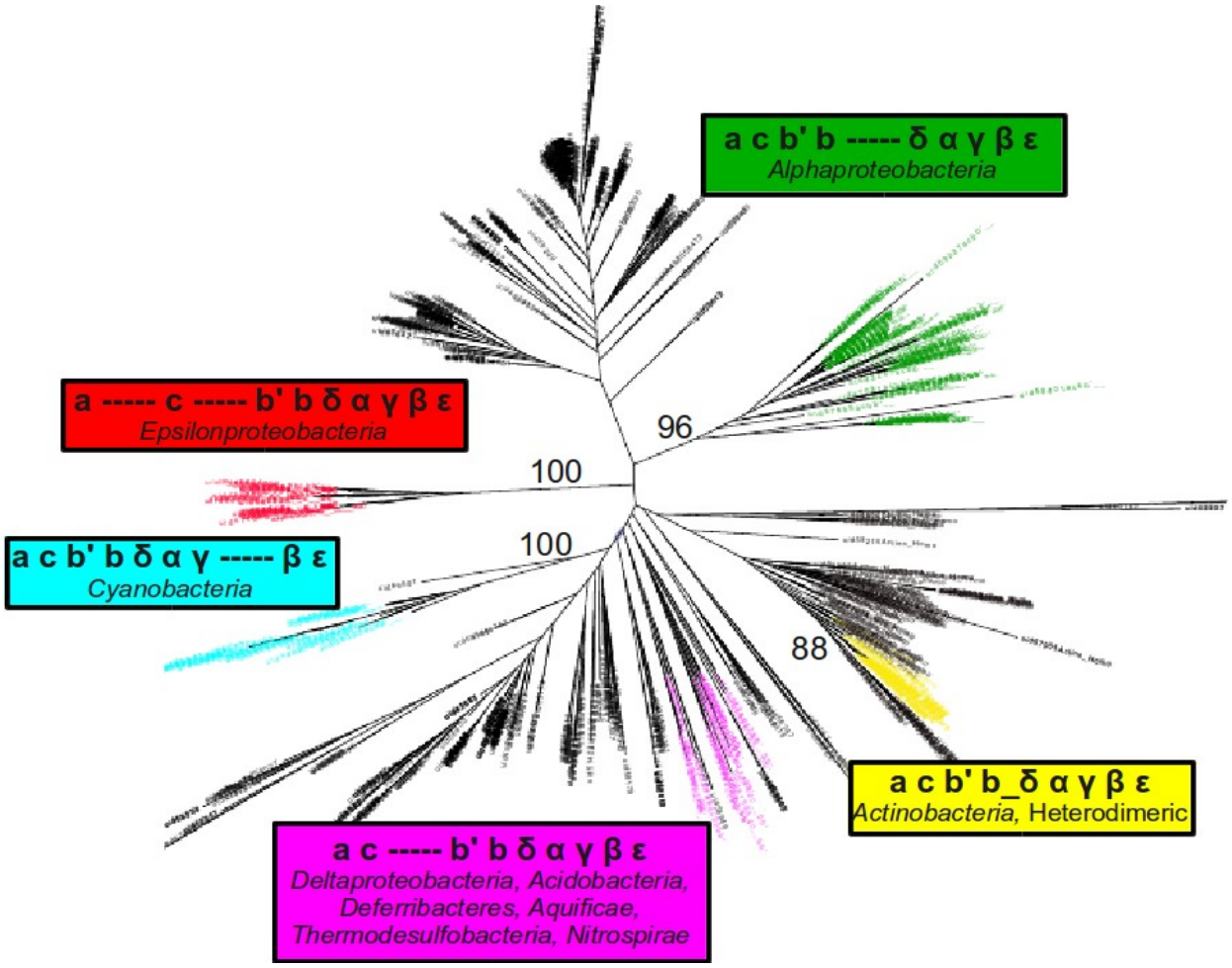

Figure 3.6: Distribution of patterns of ATP synthase transcriptional units for heterodimeric systems on the phylogenetic tree of *Bacteria*. 20 proteins from the large subunits of ribosomal proteins (L1, L2, ..., L36) were extracted for each organism and concatenated together and used to make a multiple sequence alignment. The alignment was then curated by Gblocks to keep conserved columns. The resulting file was then used by RAxML to make the phylogeny. The resultant tree has been highlighted based on the different patterns found in this study. Coloured branches correspond to heterodimeric groups with the discovered pattern shown in each box. Numbers provide bootstrap values for each major branch. The numbers for magenta branch are very low and for this reason this clustering is not very well supported. In the *Actinobacteria* section there are two outliers without fusion of the second *b* to the delta but they are not visible.

ancestor was homodimeric, this tree shows at least four cases (green, red, yellow, and blue clades) of independent clustering where the operon pattern is seen in only one class of *Bacteria* (with bootstrap values bigger than 70; values of 70% or higher are likely to indicate reliable groupings (28)). There is an exception which is the "ac-----b'bδαγβε" pattern (magenta) present in several related taxa (shown in the box), separated by low bootstrap values. It is therefore uncertain whether or not single gene duplication and gene rearrangement events are responsible for this set. Another observation is that among all phyla of *Bacteria, Actinobacteria* is the only phylum with both homo and heterodimeric organisms, regarding the stator stalk of ATP synthase. The heterodimeric group (shown in yellow), mostly *Mycobacteria*, has a fusion of *b* and delta which makes this group one of the exceptions of heterodimeric stator stalk without splitting of the operon. In general the tree confirms the usefulness of the gene structure patterns for further studies. A summary of patterns and number of organisms in each category is presented in Figure 3.4.

3.3.4 Phylogeny of the heterodimers based on the *b/b'* sequences.

In order to further study the evolution of *b/b'* sequences and their similarities it was decided to evaluate their phylogeny. Since there are two sequences for the *b* subunit in these organisms (*b* and *b'*) and we are not sure about their assignment, it was decided to call the translated protein form the first gene (in the direction of transcription unit) "First" and the other one "Second". Then all the first and second protein sequences were aligned using MUSCLE and curated (columns with gaps were removed) and finally by using two different programs, RAxML and PhyML, phylogenies were made based on maximum

likelihood approach. The result of the two softwares are very similar. Figure 3.7 shows the phylogeny made by PhyML. In this figure branches were coloured based on gene structure categories. Different shades of colours were used to distinguish between "First" and "Second" proteins among groups of similar operon structure. In fact different shades of a colour show the same operon structure in each branch. Overall, there are three distinct branches for these heterodimeric organisms. In each branch all of the "First" and "Second" proteins are similar to one another. One of the branches has three groups of bacteria in it yet all first and all second sequences are similar to one another and cluster together.

As mentioned earlier the length of the *b* subunit has been used as a criterion for the assignment of *b/b'* proteins. It was important to see how these different groups arrange based on the length of the proteins. So, the same tree was coloured this time based on the length. To compare these proteins based on the length one additional step was needed. Some of these sequences have an extra N-terminal region in their sequence which gets cleaved *in vivo* before assembly to the ATP synthase enzyme complex (1). To account for this, after multiple sequence alignment, the extra region preceding the membrane domain in the alignment were chopped off and then these sequences were used for comparison. The *b* and *b'* were compared to one another and the shorter and longer one was determined in each bacterium. As seen in the Figure 3.8, almost all of the "First" proteins are shorter than the "Second" ones. Even in the branch containing three groups of operon structures all cases of "First" are shorter than the "Second" proteins. *Candidatus*

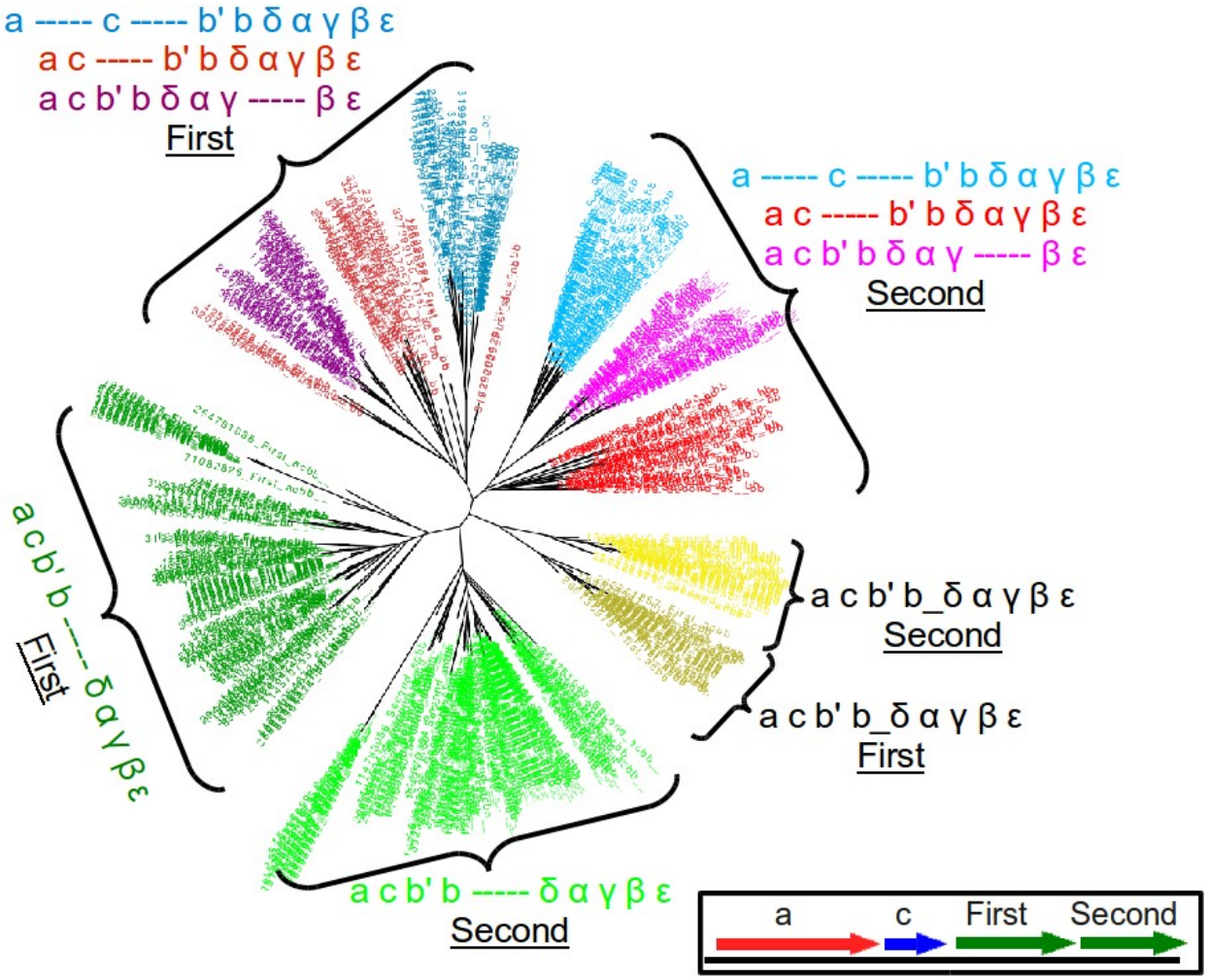

Figure 3.7: Phylogeny of all heterodimeric *b/b'* sequences highlighted based on the found patterns. Sequences of *b* and *b'* of all heterodimeric organisms were added together and used to make a multiple sequence alignment file. In the alignment file columns with gaps were removed and then the resultant curated file was used by PhyML to make the phylogeny based on maximum likelihood. The tree was highlighted based on different patterns found in this study. Because this tree is the phylogeny of all heterodimers together, in each heterodimeric organism there are two types of *b* subunit available. Since we are not sure about the nomenclature of these sequences the protein encoded by the first gene (in the direction of translation) was called "First" and the protein encoded by the other gene was called "Second" (two green arrows in the box below the figure). In each group (gene structure pattern) different shades of similar colours was used to distinguish among "First" and "Second" proteins.

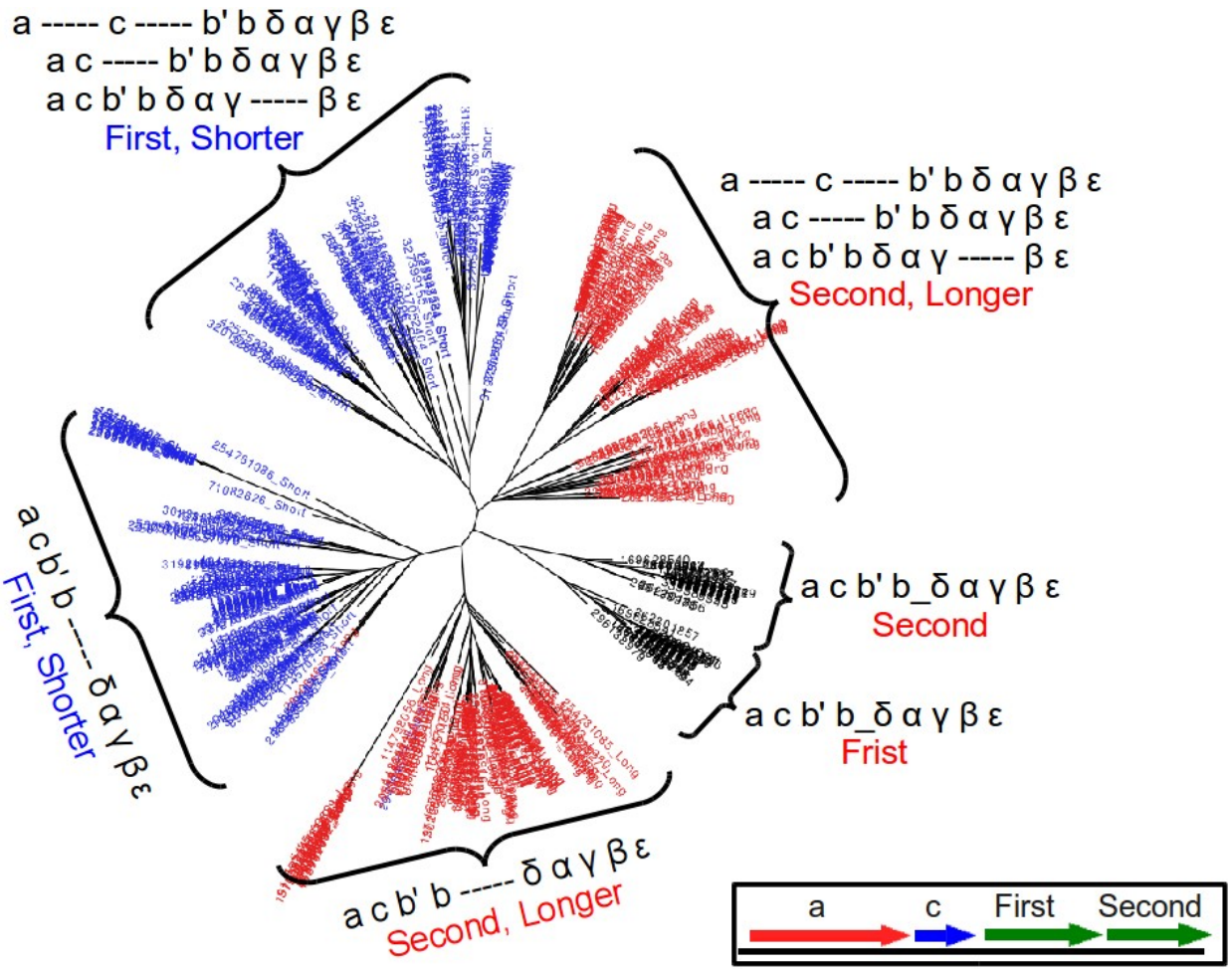

Figure 3.8: Phylogeny of all heterodimeric *b/b'* sequences highlighted based on their length. The previous tree (Figure 3.7) was highlighted this time based on the length of *b/b'* pairs in each organism. Sequences of *b* and *b'* in each organism were aligned and the extra N-terminal region of them was removed. The curated sequences in each organism compared and the "longer" and "shorter" one of each pair was determined. In this tree shorter sequences are coloured blue and longer ones red. Since there is a fusion between the second *b* protein and the delta in *Actinobacteria* (not possible to compare that to the first *b* protein), that branch on the tree was not coloured. Since we are not sure about the nomenclature of the *b* subunits the protein encoded by the first gene (in the direction of translation) was called "First" and the protein encoded by the other gene was called "Second" (two green arrows in the box below the figure).

Puniceispirillum marinum, a member of *Alphaproteobacteria*, is the only organism in which the "First" protein is longer than the "Second" protein.

3.3.5 Phylogeny of all homodimeric and heterodimeric *b* and *b'*.

To see similarities among all sequences together we looked at the phylogeny of all homo and heterodimers together at a same time. To do this all homodimeric and heterodimeric sequences were aligned using MUSCLE and columns with gaps were removed (curation). Phylogenies were made using PhyML using maximum likelihood approach. Figure 3.9 shows the result of PhyML program. Then the tree was coloured again based on gene structures, similar to the Figure 3.7. In this tree again three groups of "First" and "Second" proteins are seen. In each one of the major branches all First and all Second proteins are clustered which shows the similarity among these proteins. It is interesting that these proteins can still cluster together similar to Figure 3.7 even when the homodimeric *b* subunits (coloured black) are present in the process. In one of the branches (the one containing two shades of yellow) the homodimeric *b* subunits of *Actinobacteria* (central black), are clustered with the two clades of heterodimeric *b* and *b'* subunits of the same phylum. This shows similarity of the homodimeric and heterodimeric *b* sequences in this phylum of *Bacteria*. In another branch, where all three groups of "Second" proteins are clustered, some homodimeric sequences are also available which can be the result of either similarity or program error (since this tree was only done by PhyML and there is no bootstrap value available for these branches).

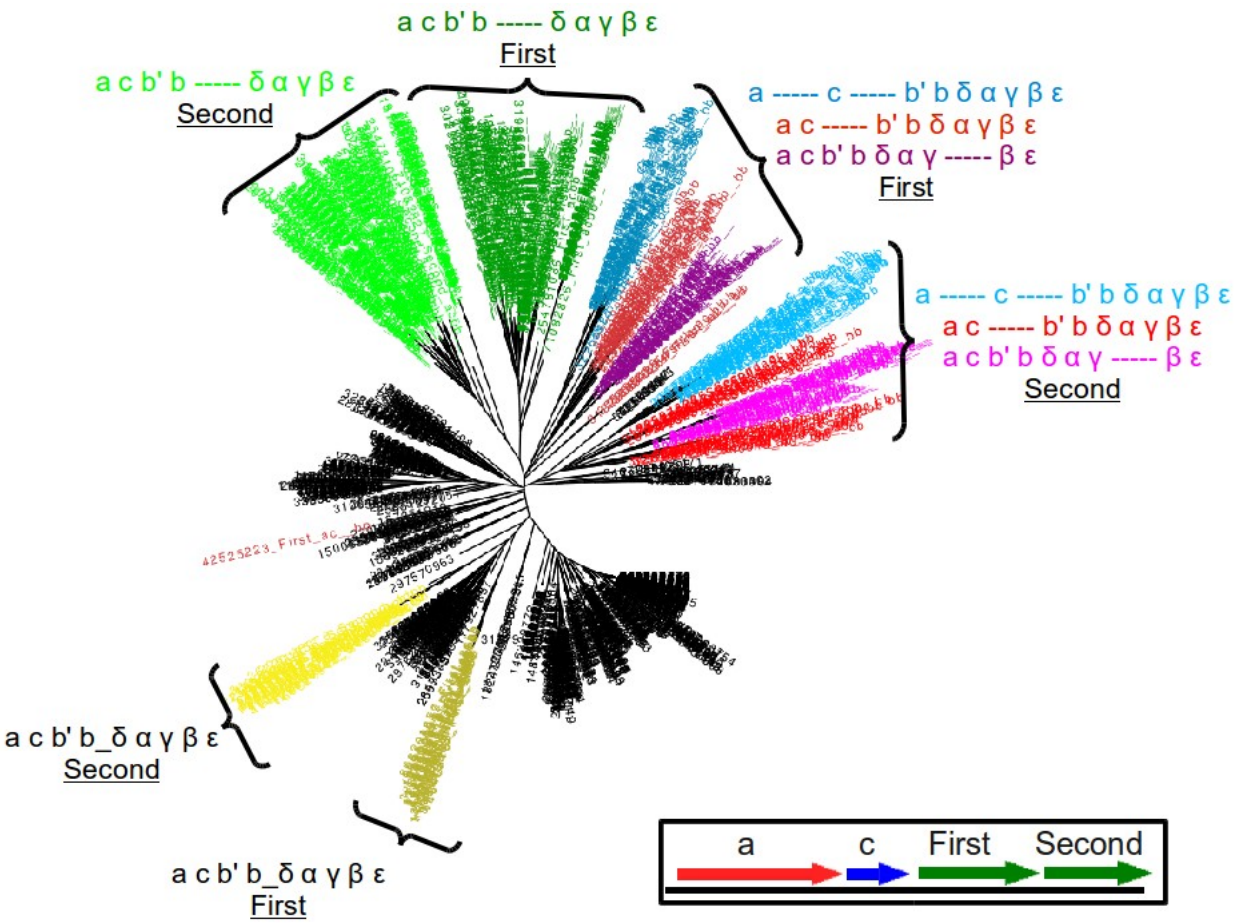

Figure 3.9: Phylogeny of all homodimeric *b* and heterodimeric *b/b'* sequences together highlighted based on the found patterns. The sequences of all homodimeric and heterodimeric *b* subunits were added together and used to make a multiple sequence alignment file. The columns containing gaps were removed from the alignment and resultant curated file was used by PhyML to make the phylogeny based on maximum-likelihood. Heterodimeric branches of the tree were coloured similar to the Figure 3.7. Different shades of colour was used to distinguish among "First" and "Second" proteins in each group. The black clades are homodimeric *b* subunits.

3.3.6 Mutual information analysis.

The mutual information study is based on the fact that positions in the sequence of a protein are not independent of one another. Mutations in one position are sometimes accompanied by other mutations in another part of the protein which has contact or interaction with the original one (29,30,31). This approach uses co-variation values between residues to show important positions and also points of contact in a sequence. This approach was started using the biggest group of heterodimeric bacteria; *Alphaproteobacteria*. For the mutual information studies it is necessary to have large number of sequences to have a good signal. In practice, more than 125 sequences are usually required to be able to analyze the data. Since the number of sequences in the completely sequenced organisms was not enough BLAST searches were done to increase them. 268 sequences were found as a pair of "First" and "Second" and used for mutual information studies. Applying similarity cut-off of 90% to the sequences reduced the number to 151 sequences and the resultant data was used for making alignments. At the moment, the result of this study is not ready yet and needs more work to be finalized in the future.

3.4 Discussion

In *E. coli* and some other species of *Bacteria*, ATP synthase has a homodimeric stator stalk, composed of two identical *b* subunits. On the other hand in chloroplasts and some other bacteria (*e.g.* photosynthetic bacteria) there are two different types of *b* subunits, *b* and *b'* (in chloroplasts called I and II), in the stator stalk. While structure of the *b* subunit in *E. coli* has been the focus of several different studies, the evolution of

this protein and its counterparts in other *Bacteria* (with heterodimeric stator stalk) and the way they arose has not been well investigated. There is one study on similarity between chloroplast subunit II and *E. coli* subunit *b* (32) but in general there is no comprehensive study regarding the relationship between the heterodimeric and homodimeric subunits and their evolution.

In *Bacteria* the two types of *b* subunits are related to *E. coli b* regarding their sequence. The main similarities are the distribution of hydrophobic N-terminal and charged region (Figure 1.3) and the 11-residue hendecad pattern (Figure 1.4) which is seen in them the same as *E. coli b*. However, when the two subunits of one organism are aligned they show significant divergence in sequence (33). This has raised the question of assignment for each one of the subunits since some differences have been observed in the interaction of these subunits with one another and also with the ATP synthase complex. One old way of assignment is using length of *b'* (or subunit II in chloroplasts) which has been reported to be shorter in Synechocystis (and spinach) (34). This idea of the *b'* being shorter than the other subunit (1) has been used as a criterion for assignment of these subunits since then without any comprehensive study on its validity. In studying the assignment of these proteins the first step to compare the *b* subunit sequences is usually by making an alignment of all known *b* or *b'* sequences. Then the resulting multiple alignment file can be used in making phylogenetic trees or BLAST search procedures to further analyze groups of similar proteins. The present study however, shows that it might be necessary to further separate the sequences and then analyze their features since

organisms with heterodimeric stator stalk are not all from the same ancestor. This is a key step. This idea started by observing that in some cases of heterodimeric organisms there is also a change in the ATP synthase operon structure (1, 9-15) which divides it into several different transcriptional units. Accordingly, instead of working directly with *b* sequences we started by looking for patterns in the *unc* operon in general. Patterns show separation points throughout evolution and we found interesting relationship between the discovered patterns and the duplication of the *b*. Applying these gene structures and extra step of classification has shown its significance according to phylogenetic studies of both homo and heterodimer sequences. It also has provided a way of classification of *Bacteria* and finding related features before working with *b* sequences.

One of the observations was the general absence of split of the ATP synthase operon in the homodimeric organisms. Only a few of the homodimers were found to have a split operon while the majority of them have one complete single operon. Exceptions are among phyla of bacteroidetes and chlorobi. Bacteroidetes have the same gene arrangement as the heterodimeric *Cyanobacteria*. *Chlorobi* are photosynthetic organisms but they do not have heterodimeric stator stalk. They have a unique pattern in their operon structure which is only seen in this phylum; "acbδ-----αγ-----βε". These exceptions were excluded from the further studies since their number was small relative to the rest of the database. In the heterodimeric groups on the other hand, there is usually a separation of the genes encoding ATP synthase subunits into two or more transcriptional units separated from each other by various lengths of DNA sequence. The exception in

this case is the phylum of *Actinobacteria* containing *Mycobacteria* species. The *b* subunit gene in these bacteria is fused to the δ gene with the help of a linker sequence which is about 100 amino acids long. For this reason they have a single operon as well. Another observation in this phylum is the existence of some bacteria with two genes for *b* and *b'* (heterodimers), but still no split in their ATP synthase operon. This study found two bacteria of this kind: *Pseudonocardia dioxanivorans CB1190* and *Segniliparus rotundus*. One other interesting observation was that the genes encoding the two *b* subunits in heterodimers are always in the same transcriptional unit, regardless of what happend to the remainder of the ATP synthase operon. On the other hand the splitting happens between subunits other than *b/b'*. This could be important since it helps to have equal amount of expression for the two *b* subunits while having them in different transcriptional units puts each one of the subunits under a different controlling mechanism that would likely produce unequal amounts of each subunit, resulting in a higher risk of homodimer formation for the more abundant subunit.

The tree of life of these *Bacteria*, based on ribosomal proteins, shows independent evolution of these operon arrangements where organisms with the same pattern usually are clustered together independent from organisms having a different pattern. This distribution of operon arrangements among *Bacteria* shows their independent origin and shows that the *b/b*' heterodimeric system developed independently from several different initial events. We propose that the duplication of *b* occurred at the same time as a split was introduced into the ATP synthase operon (we do not know which one happened first)

which remained in that specific class of *Bacteria* and this process happened several times independent from the others. Why there is a split in the operon in the heterodimeric organisms is still unknown to us.

Some branches of the tree of *Bacteria* are worth mentioning separately because of strange clustering of their members. The pattern "ac-----b'bδαγβε" (magenta in the Figure 3.6) was observed in more than one branch in the tree of *Bacteria* and is seen in bacteria belonging to taxa *Acidobacteria, Deltaproteobacteria, Deferribacteres, Aquificae, Thermodesulfobacteria, And Nitrospirae*. Analyzing the bootstrap values for each one of these separate branches for one pattern, showed that other than the clade of *Deltaproteobacteria*, other branches are not very well supported and for this reason we can not tell for sure that they are positioned correctly. What we can conclude is the independent clustering of this group in *Bacteria* compared to others although we do not know if it was only one event or several. *Actinobacteria* is another interesting branch in the tree of *Bacteria* (yellow clade in the Figure 3.6 represents the heterodimeric ones) where both homodimers and heterodimers are present. Homodimers and heterodimers in this group cluster together independently. Also, the two heterodimeric organisms without fusion between *b* and delta, cluster with homodimeric *Actinobacteria*. Analyzing their bootstrap value shows insignificant numbers which can be an inaccuracy in the phylogeny inference. Regardless of this, heterodimerization in this branch (yellow in the Figure 3.6) counts as one independent clustering event independent from other events.

Phylogeny based on the *b* subunits among all heterodimeric sequences together

also showed three different ways that the *b/b'* arrangement originated (three major branches radiated from the central point in the Figure 3.7) with all "First" and "Second" proteins in each bacterium clustering together independent from the rest. In fact, in every branch only "First" or "Second" proteins are present without any exception. Analyzing the same tree with the length of each "First" protein compared to the "Second" protein showed one case of exception. In other cases the "First" protein is shorter than the "Second" ones (Figure 3.8). Further analysis of the phylogeny of the heterodimeric *b* subunits of *Alphaproteobacterial* group with more sequences included revealed more cases of the "First" protein being longer than the "Second" ones (data not shown). Accordingly, an assignment based on the position of the subunits in the operon (*i.e.* "First" or "Second") seems to be more reliable than the length alone. The same result was made when the tree of all homodimeric and heterodimeric *b* subunits was made together. Again, major branches of the tree containing only "First" or "Second" proteins are observable (Figure 3.9). This observation is important because it further emphasizes the fact that there is similarity between sequences in each group (either "First" or "Second" proteins) even when all homodimeric and heterodimeric sequences are present in the group analyzed. This can be very helpful when each one of these groups of similar sequences are represented as a PSSM for database searching. In the future this can be used to find similar sequences more accurately.

The fact that there are several different ways of *b/b'* divergence is very helpful for future studies of heterodimeric systems. It implies that it is not appropriate to assume all

b or *b'* sequences are orthologous to each other. It depends on the clade they belong to. The subunits in the different clades will instead be paralogous to one another. This is very important when searching for new related sequences. Representing each one of the groups (for example all of the "First" proteins in the *Alphaproteobacteria*) by a specific PSSM may help find related sequences more correctly. This further emphasizes the importance of classification of *Bacteria* based on the discovered patterns. The discovered patterns may also be helpful in mutual information analysis. This analysis may reveal the *b* subunit contact areas in the dimer (both inter and intra-molecular interactions) which can be helpful in the study of offset model. It also can confirm the possible hendecad pattern in the *b* subunit sequence. In the past this analysis always has failed in the lab because of the unfamiliarity with different ways of *b/b'* speciation. We think it will reveal much more information this time with the present knowledge about heterodimeric systems.

At this point the evolutionary advantage of having a heterodimeric system instead of having a homodimeric stator stalk of identical *b* subunits is not clear. Our study shows that there are several ways of developing *b/b'* systems among *Bacteria* with a coinciding split in the ATP synthase operon but it does not show any advantage of having a heterodimeric system. We know that in *Rhodobacter capsulatus* (35) and *Aquifex aeolicus* (36) there is one *b* and one *b'* incorporated in each ATP synthase enzyme molecule, presumably playing different roles based on our offset model and proximity of only one of them to delta subunit (3). Also we know that b^N and b^C subunits in *E. coli* are

not only asymmetric with respect to one another but also have different interactions with the enzyme (3). But, there is lack of experimental evidence showing the obligate heterodimer formation in these bacteria. For this reason it is not possible to conclude that speciation to a more complex system, heterodimeric *b/b'* in this case, is necessarily evolutionarily advantageous. Recently, a study showed that not all complexities that evolve in a system are necessarily improvements. There is a theory, called constructive neutral co-evolution, which explains how neutral processes might drive a system towards complexity (37) without any benefit for the function of the system. At this point what this study shows is that several different, and independent speciation events happened for the *b/b'* system to develop and the classification of organisms based on these events coincides very well with the evolutionary tree of *Bacteria*. This classification can be used as a basis for further studying structural features of *b* and *b'* proteins however the functional characteristics of these proteins need to be studied in more detail to reveal possible advantages for having a heterodimeric system.

3.5 References

1 Cozens AL, Walker JE. (1987) The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium Synechococcus 6301. Support for an endosymbiotic origin of chloroplasts. *J Mol Biol.* **194**, 359-83.

2 Del Rizzo PA, Bi Y, Dunn SD. (2006) ATP synthase b subunit dimerization domain: a right-handed coiled coil with offset helices. *J Mol Biol.* **364**, 735-46.

3 Wood KS, Dunn SD. (2007) Role of the asymmetry of the homodimeric b2 stator stalk in the interaction with the F1 sector of Escherichia coli ATP synthase. *J Biol Chem.* **282**, 31920-7.

4 Tiburzy HJ, Berzborn RJ. (1997) Subunit II (b') and not subunit I (b) of photosynthetic ATP synthases is equivalent to subunit b of the ATP synthases from nonphotosynthetic eubacteria. Evidence for a new assignment of b-type F0 subunits. *Z Naturforsch C.* **52**, 789-98.

5 Poetsch A, Berzborn RJ, Heberle J, Link TA, Dencher NA, Seelert H. (2007) Biophysics and bioinformatics reveal structural differences of the two peripheral stalk subunits in chloroplast ATP synthase. *J Biochem.* **141**, 411-20.

6 Walker JE, Saraste M, Gay NJ. (1984) The unc operon. Nucleotide sequence, regulation and structure of ATP-synthase. *Biochim Biophys Acta.* **768**, 164-200.

7 Ivey DM, Krulwich TA. (1991) Organization and nucleotide sequence of the atp genes encoding the ATP synthase from alkaliphilic Bacillus firmus OF4. *Mol Gen Genet.* **229**, 292-300.

8 Ohta S, Yohda M, Ishizuka M, Hirata H, Hamamoto T, Otawara-Hamamoto Y, Matsuda K, Kagawa Y. (1988) Sequence and over-expression of subunits of adenosine triphosphate synthase in thermophilic bacterium PS3. *Biochim Biophys Acta.* **933**, 141-55.

9 Tybulewicz VL, Falk G, Walker JE. (1984) Rhodopseudomonas blastica atp operon. Nucleotide sequence and transcription. *J Mol Biol.* **179**, 185-214.

10 Falk G, Hampe A, Walker JE. (1985) Nucleotide sequence of the Rhodospirillum rubrum atp operon. *Biochem J.* **228**, 391-407.

11 Borghese R, Turina P, Lambertini L, Melandri BA. (1998) The atpIBEXF operon coding for the F0 sector of the ATP synthase from the purple nonsulfur photosynthetic bacterium Rhodobacter capsulatus. *Arch Microbiol.* **170**, 385-8.

12 Borghese R, Crimi M, Fava L, Melandri BA. (1998) The ATP synthase atpHAGDC (F1) operon from Rhodobacter capsulatus. *J Bacteriol.* **180**, 416-21.

13 H S Van Walraven, R Lutter, J E Walker. (1993) Organization and sequences of genes for the subunits of ATP synthase in the thermophilic cyanobacterium Synechococcus 6716. *Biochem J.* **294**, 239-251.

14 Lill H, Nelson N. (1991) The atp1 and atp2 operons of the cyanobacterium Synechocystis sp. PCC 6803. *Plant Mol Biol.* **17**, 641-52.

15 Curtis SE. (1987) Genes encoding the beta and epsilon subunits of the proton-translocating ATPase from Anabaena sp. strain PCC 7120. *J Bacteriol.* **169**, 80-6.

16 Stormo GD, Schneider TD, Gold L, Ehrenfeucht A. (1982) Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli. *Nucleic Acids Res.* **10**, 2997-3011.

17 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *J Mol Biol.* **215**, 403-10.

18 Dickson RJ, Wahl LM, Fernandes AD, Gloor GB. (2010) Identifying and seeing beyond multiple sequence alignment errors using intra-molecular protein covariation. *PLoS One*. **5**, 11082.

19 Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.* **25**, 1422-3.

20 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-402.

21 Pritchard L, White JA, Birch PR, Toth IK. (2006) GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics.* **22**, 616-7.

22 Edgar RC. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-7.

23 Talavera G, Castresana J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst*

Biol. **56**, 564-77.

24 Guindon S, Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* **52**, 696-704.

25 Stamatakis A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* **22**, 2688-90.

26 Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science.* **311**, 1283-7.

27 Yutin N, Puigbò P, Koonin EV, Wolf YI (2012) Phylogenomics of Prokaryotic Ribosomal Proteins. *PLoS ONE* **7**, 36972.

28 Hillis, D.M. and Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analyses. *Syst. Biol*. **42**, 182-192

29 Fitch WM, Markowitz E. (1970) An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem Genet*. **4**, 579-93.

30 Yanofsky C, Horn V, Thorpe D. (1964) Protein structure relationships revealed by mutational analysis. *Science*. **146**, 1593-4.

31 Pazos F, Valencia A (2008) Protein co-evolution, co-adaptation and interactions. *EMBO J* **27**, 2648-2655.

32 Tiburzy HJ, Berzborn RJ. (1997) Subunit II (b') and not subunit I (b) of photosynthetic ATP synthases is equivalent to subunit b of the ATP synthases from nonphotosynthetic eubacteria. Evidence for a new assignment of b-type F0 subunits. *Z Naturforsch C*. **52**, 789-98.

33 Dunn, S. D., Cipriano, D. J., and Del Rizzo, P. A. (2004) In *Handbook of ATPases* (Futai, M. and Wada, Y., eds), pp. 311-318, Springer-Verlag, Weinheim.

34 Weber J. (2006) ATP synthase: subunit-subunit interactions in the stator stalk. *Biochim Biophys Acta*. **1757**, 1162-70.

35 Gabellinia N, Gaoa Z, Eckerskorna C, Lottspeicha F, Oesterhelta D. (1988) Purification of the H+-ATPase from Rhodobacter capsulatus, identification of the F1F0 components and reconstitution of the active enzyme. *Biochim Biophys Acta - Bioenergetics.* **934**, 227-234.

36 Peng G, Bostina M, Radermacher M, Rais I, Karas M, Michel H. (2006) Biochemical and electron microscopic characterization of the F1F0 ATP synthase from the hyperthermophilic eubacterium Aquifex aeolicus. *FEBS Lett.* **580**, 5934-40.

37 Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. (2012) Evolution of increased complexity in a molecular machine. *Nature*. **481**, 360-4.

Chapter 4: Final conclusions

4.1 Conclusion

F_1F_o Adenosine triphosphate (ATP) synthases are the major producer of universal energy currency, ATP, in living organisms in both oxidative phosphorylation and photophosphorylation processes and they are located in the cytoplasmic membranes of bacteria, inner membranes of mitochondria, and the thylakoid membranes of chloroplasts. They all function by a rotary mechanism in which the electrochemical energy of a proton gradient is used to produce ATP from ADP and inorganic phosphase. F_1F_o ATP synthases are composed of two linked sectors: the membrane-integral F_o sector which catalyzes the movement of protons across the membrane and the membrane-peripheral F_1 sector which catalyzes the synthesis or hydrolysis of ATP. The enzyme's peripheral stalk, composed of the $b_2\delta$ subunits, is the major link between the two sectors. It serves as the stator that holds the F_1 sector and its catalytic sites against the movement of the rotor. In *E. coli*, the *b* component of the peripheral stalk is a homodimer of identical *b* subunits. The *b* subunit resides in the membrane by its N-terminal part and stretches the length of F_1F_o complex, making contact with the δ subunit at the top of F_1 sector (1,2).

Peripheral stalks in *Bacteria* are not always homodimeric. Photosynthetic bacteria such as *Cyanobacteria*, chloroplasts, and some other species among *Proteobacteria* assemble a heterodimer of *b*-type subunits called *b* and *b'* within their ATP synthase. Both subunits have been shown to be present in ATP synthase from some species including *Rhodobacter capsulatus* (3) and *Aquifex aeolicus* (4). There are some studies that support

the idea that heterodimers of *b/b'* form preferentially. Sedimentation and chemical cross-linking studies showed that the cytoplasmic domains of the *b* and *b′* subunits of *Synechocystis* form a highly extended heterodimer. These data support the idea that *b* and *b'* subunits in ATP synthases of mentioned organisms exist in the complex in a heterodimeric form, occupying positions and performing functions highly similar to those of the homodimeric *b* subunits of *Bacteria* like *E. coli* (5).

The *b* subunit of *E. coli* consists of 156 residues and has a molecular weight of 17264 Da. Based on deletion studies it has been divided into four domains. The insoluble N-terminal part of *b* subunit which resides in the membrane, is called transmembrane domain ($b_{M1\text{-}I23}$). The soluble portion consists of the tether domain ($b_{E24\text{-}A53}$) which is the segment between the surface of the membrane and the bottom of F_1 sector, the dimerization domain ($b_{T53\text{-}K122}$) which extends up to one of the αβ interfaces in F_1, and at the end, the δ binding domain which is the C-terminal of *b* subunit and is required for the interaction of the peripheral stalk and F_1 ($b_{Q123\text{-}L156}$) (1,2).

Although there is no complete high-resolution structure available for the *b* dimer, some parts of it has been studied separately. The isolated dimerization domain of the *b* subunit has been characterized as an atypical, parallel, two-stranded coiled coil and sequence analyses of this region have identified an 11-residue hendecad pattern, with positions denoted *abcdefghijk* (6). Hendecad patterns are indicative of right-handed coiled coils in which 11 residues make three turns of the helix relative to the interhelical axis (7). Del Rizzo *et al.* (8) confirmed the assignment of *abcdefghijk* by introducing

cysteine mutations in these positions and assessment of the stabilities of disulfide-linked dimers. Also, the study showed that the two *b* subunits are offset by 5.5 residues in the dimerization domain and one of the subunits is shifted toward the N-terminal and the other toward the C-terminal. The C-terminally shifted subunit was labeled b^C and the other subunit b^N. Other experiments have indicated specific roles for each one of the subunits. For example it has been shown that the subunit which is closer to δ and is more important for F_1 binding is the b^N (9). These data suggest that the two subunits are not only asymmetric with respect to each other but also have different interaction with the enzyme complex. Recently, this unusual coiled coil structure which only was suggested based on experimental and bioinformatics investigations has found support from recent crystallographic analysis of the related heterodimeric complex from the *Thermus thermophilus* A-type ATP synthase (10).

The focus of this project was to study the structure, function, and interactions of the *b* subunit of *E. coli* by using protein chemistry and bioinformatics techniques. In work described in Chapter 2, we developed an obligate heterodimeric system of *b* subunits in *E. coli*. As the stator stalk of *E. coli* is a homodimer of identical *b* subunits, it is difficult to design experiments to study its structure and interactions. Chimeric approach was used for differentiating between the two subunits and understanding their location and orientation and any specific occupation of the b^N or b^C positions. Chimeric approach has been used previously with little achievements (11). Formation of homodimeric stator stalks as well as heterodimeric ones and difficulty in expression of the chimeric subunits

in the cell (due to the limitation in the number of plasmids transformed) were some of the shortcomings of the previous work. Here, we replaced residues 34-110 of *E. coli b* subunit with corresponding residues of *b* and *b'* sequences of *Rhodobacter capsulatus*, a mesophile, instead of *Thermosynechococcus elongatus*, and we showed that the newly developed chimeric system is an obligate heterodimeric system. Only cells with heterodimeric chimera were able to grow on succinate. No homodimer formation was observed. In addition, the chimera is designed in such a way that the two subunits are next to each other in one plasmid containing all *uncF* genes. Accordingly, it is not only possible to study the *b* subunits itself, but also the interaction of these chimeric subunits with the rest of ATP synthase subunits. The expression of these two chimeric subunits were linked together by using a ribosomal re-initiation site between the two genes. Specific detection of each chimeric subunit was also made possible by adding sequences of V5 and His tag to *b'* and *b* chimeric subunits respectively. Our data showed that the same level of protein was made in cells for the two subunits based on the darkness of the bands in Western blot analysis. The incorporation of the two subunits into the ATP synthase was also the same for both chimeric subunits as was seen by ATP synthase fractionation. Finally, cysteine cross-linking studies showed that the orientation of the *b* and *b'* subunits in the stator stalk was the same as in the wild type *R. capsulatus*, in contrast to the previous work. Also, it confirmed the incorporation ratio of 1:1 to the ATP synthase enzyme for the two chimeric subunits.

When the chimeric system was set-up, we started to study some of the questions

regarding the structure and features of the stator stalk of the *E. coli* ATP synthase. A series of cysteine mutations were made in the region between residues 2-10 to confirm the offset model in the transmembrane domain. The system proved to be very easy and simple to work with by providing fast and independent introduction of cysteine mutations in desired positions. Unfortunately, for reasons not well understood, mutated constructs failed to complement cells on succinate, except for two cases. Also, no disulfide formation was observed for any of the combination of cysteine mutations. Experiments involving single cysteine mutations in the mentioned positions would be helpful to see what positions are prone to malfunction when mutated. In addition, more work should be done to be sure about the amino acids in the helix-helix interfaces.

Another application of the system which was started was in the determination of the role of each one of the subunits in the stiffness or flexibility of the stator stalk of *E. coli* by introducing glycine mutaions. A series of one to five glycine mutations were introduced in each one of the chimeric subunits (positions 79-83) independently from the other one. Glycine mutations make the helix less stiff without having dramatic effect on the helix structure. The preliminary growth test on succinate showed that stiffness of the helices of each subunit was important since introducing more than two glycine mutations disrupted the functionality. However, better growth was observed in the set with glycine mutations in the *b* subunit compared to mutations in the *b'*. Future work involving studying proton pumping activities of these constructs and also combining glycine mutations in the two subunits will help us understand the role of these subunits in the

stiffness of the stator stalk in *E. coli*.

In part two, we sought the evolution and assignment of the *b* and *b'* subunits through bioinformatic methods. It should be mentioned that *Cyanobacteria* were the first organisms investigated for their heterodimeric *b* subunits and in those cases the length criterion was used for the assignment of *b* and *b'*;this convention also has been used today. However, considering the different interactions of the two subunits with one another and the different interactions that they have with the rest of the ATP synthase enzyme, using length as the only criterion seems to be arbitrary. We tried to study the subunits based on sequence similarity and phylogenetic characteristics.

Inspired by some limited previously studied cases among *Cyanobacteria* and *Rhodobacteria*, we decided to look for possible patterns of gene organization in the ATP synthase operon. We found that there is an interesting connection between the gene duplication of the *b* subunit and the structure of the ATP synthase operon. Usually, when gene duplication happened for the *b* subunit also a split occurred in the operon which broke it into several different transcriptional units, separated from each other by varying lengths of DNA. In this study, we were able to categorize heterodimeric organisms into five different groups based on these ATP synthase gene structure. Putting these heterodimers onto the tree of *Bacteria*, determined by ribosomal protein sequences, revealed the independent evolution of these patterns among heterodimeric organisms and provided evidence for the number of times the gene duplication happened among *Bacteria*. Our data showed at least three independent gene duplication events for the *b*

subunit. Following that, the relationship between subunits, defined by gene order, was determined by doing phylogenetic studies based on *b* subunit sequences. This led to the conclusion that in each group of organisms, related by gene organization, there is similarity between all proteins that are encoded by the first gene in the operon, as they cluster together in the tree. This is also the case for the proteins that are encoded by the second gene. Further studies of these trees, by looking at the length of the sequences, suggested that considering the length as the only criterion for the assignment is not sufficiently complete. In some cases there are exceptions where the protein encoded by either the first gene or the second gene is not necessarily shorter or longer respectively.

Being able to categorize heterodimeric organisms based on ATP synthase gene organization seems to be promising for future studies. Representing each group by a PSSM and BLAST searching the database using these PSSMs will help to find related sequences. Another study which started in this thesis was the mutual information studies among the *b* subunits. The goal of this kind of study was to determine points in the sequence of a protein that are contact regions, either inter- or intra-molecular. The importance of this study for the *b* subunit is where the result of this approach helps to discover sites where key subunit-subunit interactions occur. The important prerequisite step for this approach is to have a sufficiently large number of related sequences, easily alignable, to be able to make a good multiple sequence alignment. All previous attempts in the lab were unsuccessful in the past because of lack of knowledge about the different ways of gene duplication among *b* subunits. This time since we have a promising way of

categorizing organisms based on common features, we believe this analysis will be successful.

4.2 References

1 Dunn SD, Revington M, Cipriano DJ, Shilton BH. (2000) The b subunit of Escherichia coli ATP synthase. *J Bioenerg Biomembr.* **32**, 347-55.

2 Claggett SB, Grabar TB, Dunn SD, Cain BD. (2007) Functional incorporation of chimeric b subunits into F_1F_0 ATP synthase. *J Bacteriol.* **189**, 5463-71.

3 Gabellinia N, Gaoa Z, Eckerskorna C, Lottspeicha F, Oesterhelta D. (1988) Purification of the H+-ATPase from Rhodobacter capsulatus, identification of the F1F0 components and reconstitution of the active enzyme. *Biochim Biophysic Acta.* **934**, 227-234.

4 Peng G, Bostina M, Radermacher M, Rais I, Karas M, Michel H. (2006) Biochemical and electron microscopic characterization of the F_1F_0 ATP synthase from the hyperthermophilic eubacterium Aquifex aeolicus. *FEBS Lett.* **580**, 5934-40.

5 Dunn SD, Kellner E, Lill H. (2001) Specific heterodimer formation by the cytoplasmic domains of the b and b' subunits of cyanobacterial ATP synthase. *Biochemistry*. **40**, 187-92.

6 Dunn, S. D., Cipriano, D. J., and Del Rizzo, P. A. (2004) In *Handbook of ATPases* (Futai, M. and Wada, Y., eds), pp. 311-318, Springer-Verlag, Weinheim.

7 Lupas A. (1996) Coiled coils: new structures and new functions. *Trends Biochem Sci.* **21**, 375-82.

8 Del Rizzo PA, Bi Y, Dunn SD. (2006) ATP synthase b subunit dimerization domain: a right-handed coiled coil with offset helices. *J Mol Biol.* **364**, 735-46.

9 Wood KS, Dunn SD. (2007) Role of the asymmetry of the homodimeric b2 stator stalk in the interaction with the F1 sector of Escherichia coli ATP synthase. *J Biol Chem.* **282**, 31920-7.

10 Lee LK, Stewart AG, Donohoe M, Bernal RA, Stock D. (2010) The structure of the peripheral stalk of Thermus thermophilus H+-ATPase/synthase. *Nat Struct Mol Biol.* **17**, 373-8.

11 Claggett SB, Grabar TB, Dunn SD, Cain BD. (2007) Functional incorporation of chimeric b subunits into F1Fo ATP synthase. *J Bacteriol.* **189**, 5463-71.

Curriculum Vitae

Name: Ardeshir Goliaei

Post-secondary Education and Degrees:

- Tehran University of Medical Sciences
  Tehran, Iran
  2000-2006 Pharm.D. (Doctorate in Pharmacy)

- The University of Western Ontario
  London, Ontario, Canada
  2010-2012 M.Sc.

Honours and Awards:

- Ministry of Health
  Full competitive scholarship
  Tehran, Iran
  2000-2006

- Western Graduate Research Scholarships (WGRS)
  London, Ontario, Canada
  2010-2012

Related Work Experience:

Teaching Assistant
The University of Western Ontario
2012

Publications:

- Payabvash S, Ghahremani MH, **Goliaei A**, Mandegary A, Shafaroodi H, Amanlou M, Dehpour AR. Nitric oxide modulates glutathione synthesis during endotoxemia. Free Radic Biol Med. 2006 Dec 15;41(12):1817-28.
- Shahverdi AR, Khoramizadeh MR, Ghahramani MH, **Golyaee A**, Attar F, Ghahraman A. Chemopreventive effect of Cousinia shulabadensis Attar & Ghahraman ethanol extract. Afr J Tradit Complement Altern Med. 2006 Aug 28;4(1):12-6.
- **Goliaei A**, Shahriari S, Ghahremani M H, Ostad S N. Cloning and production technology of recombinant human tissue plasminogen activator. Iranian Patent No:35155. Issued May 5, 2006.
- **Goliaei A**, Keramati A R, Azizi E, Ostad S N, Ghahremani M H. The expression level of tumor suppressors $p16^{INK4a}$ and $p18^{INK4c}$ in cancer cell line using RT-PCR. 10^{th} International Pharmaceutical Sciences Conference, Tehran, Iran, 2006.
- Keramati A R, **Goliaei A**, Ghahremani M H. The profile expression of XIAP and

ERK and AKT activation following serum stimulation in serum starved HT-29 colorectal cancer cells. 10th International Pharmaceutical Sciences Conference, Tehran, Iran, 2006.